Human Factors of Integrating Speech and Manual Input Devices:

The Case of Computer Aided Design

by

Halimahtun Mohd. Khalid

University College London

Submitted for the Degree of Doctor of Philosophy

in the University of London

November 1989

ABSTRACT

The thesis investigates integrating the use of speech input and manual input devices in human-computer systems. The domain of computer aided design (CAD) is used as a case study.

A methodology for empirical evaluation of CAD systems is presented. The methodology is based on a framework that describes the input/output processes presumed to underlie performance in design activities, using behaviour protocols and performance indices as data. For modelling system behaviour, a framework derived from the Blackboard architecture of design is described. The framework employs knowledge sources to represent different behaviour types recruited during CAD performance. Variability in user behaviour throughout the investigation is explained with reference to the model.

The problems that expert CAD users experience in using manual input devices are first documented in an observational study conducted at their workplace. This demonstrates that the unitary use of manual input resulted in non-optimal behaviour. Possible solutions to these problems, using speech input for some command and data entry tasks, are explored in three experiments. In each experiment, a comparative analysis of alternative systems is made using data obtained from naive and novice users.

In Experiment 1, the use of speech as a unitary solution to the problems of manual input was also found to result in non-optimal behaviour and performance. The solution explored in Experiment 2 was to allocate some commands and alphanumeric data to each input device, using the frequency of use principle. This approach, however, entailed the additional problem of remembering which device to use. Experiment 3 evaluated the separate allocation of commands to speech input and numeric plus graphical data to manual input. Additionally, performance aids and feedback facilities were provided to users. This clear-cut assignment of device to task characteristics and the use of such aids led to an enhancement in speech performance, in addition to improving behaviour.

The findings from this research are used to develop guidelines for an integrated CAD system involving speech and manual input. The guidelines, which are intended for use by end users, CAD implementors and system designers, were validated in the workplace by the latter. Lastly, the thesis contextualises the research within an ergonomics framework, mapping the research development from problem specification to application and synthesis. Problems with the investigation are also discussed, and suggestions made as to how these might be resolved.

CONTENTS

Title Page			1
Abstract			2
Contents			3
List of Figures			7
List of Tables			8
Acknowledgements			9
CHAPTER 1.	**Introducing the Research Problem, Solutions and Application Domain**		11
	Overview of the thesis		12
	1.1	Introduction	13
	1.2	Human factors concern	14
	1.3	The problem	15
	1.4	Potential human factors solutions	19
	1.5	The application domain	22
	1.6	The research process	23
	1.7	Organisation of the thesis	25
	1.8	Summary	25
	Next chapter highlights		26
CHAPTER 2.	**A Review of Human Factors Issues of Speech and Manual Input**		27
	Overview		28
	2.1	Introduction	28
	2.2	Background to speech recognition	29
	2.3	Some ergonomic aspects of ASR systems	30
	2.4	Sources of error in ASR use	31
	2.5	Background to manual input	35
	2.6	Some ergonomic aspects of manual input devices	36
	2.7	Review of studies on speech and manual input	39
	2.8	Summary	45
	Next chapter highlights		46
CHAPTER 3.	**Background to CAD Technology and the Design Process**		47
	Overview		48
	3.1	Introduction	48

3.2 What is design and the design process? 48
3.3 What is CAD? 52
3.4 CAD/CAM development 56
3.5 CAD systems 61
3.6 Human factors in CAD 66
3.7 Empirical studies on human factors in CAD 69
3.8 Summary 71
Next chapter highlights 71

CHAPTER 4. A Behaviour-based Methodology for Empirical Evaluation of CAD Systems 72
Overview 73
4.1 Introduction 73
4.2 System assessment 76
4.3 Determining optimal design behaviour: CAD versus traditional draughting 77
4.4 Behaviour protocol analysis in design 78
4.5 Behaviour and performance measurement 84
4.6 System behaviour modelling 89
4.7 Human factors guidelines 90
4.8 Summary 90
Next chapter highlights 90

CHAPTER 5. A Blackboard Framework for Modelling System Behaviour 91
Overview 92
5.1 Introduction 92
5.2 Modelling behaviour and performance 93
5.3 The framework for the model - the blackboard architecture 95
5.4 Example blackboard systems: Whitefield's blackboard model of design 98
5.5 Blackboard architecture for modelling system behaviour 100
5.6 Constructing a blackboard model of system behaviour 107
5.7 Notional mapping of design model to system behaviour model 110
5.8 Summary 110
Next chapter highlights 112

CHAPTER 6. Study 1: Observing CAD Experts at Work - Documenting the Problems of Manual Input 113
Overview 114
6.1 Introduction 114

	6.2	Method	115
	6.3	Results	118
	6.4	Discussion	132
	6.5	Summary	136
	Next chapter highlights		136
CHAPTER 7.	**Study 2: Optimising the Performance of a Demonstrator System**		137
	Overview		138
	7.1	Introduction	138
	7.2	Description of demonstrator system	139
	7.3	Optmisation method	145
	7.4	Analysis and discussion	147
	7.5	Device assessment and recommendations	152
	7.6	Summary	154
	Next chapter highlights		154
CHAPTER 8.	**Experiment 1: Assessment of Speech Input as a Unitary Solution to the Problems of Manual Input - Comparisons of Speech Recogniser and Graphics Tablet**		155
	Overview		156
	8.1	Introduction	156
	8.2	Method	158
	8.3	Data analysis	164
	8.4	Results	164
	8.5	Discussion	180
	8.6	Summary	186
	Next chapter highlights		186
CHAPTER 9.	**Experiment 2: Assessment of Integrating Speech and Manual Input as a Potential Solution to the Problems of Unitary Speech Input - Comparisons of Device-Task Allocation Strategies**		187
	Overview		188
	9.1	Introduction	188
	9.2	Method	192
	9.3	Data analysis	195
	9.4	Results	198
	9.5	Discussion	211
	9.6	Summary	216
	Next chapter highlights		216

CHAPTER 10.	**Experiment 3: Assessment of Integrated Speech-Manual Input System as a Better Hybrid Solution - Comparisons of Performance Aids and Feedback Mechanisms**	**217**
	Overview	218
	10.1 Introduction	218
	10.2 Method	223
	10.3 Data analysis	226
	10.4 Results	228
	10.5 Discussion	239
	10.6 Summary	243
	Next chapter highlights	243
CHAPTER 11.	**Developing and Validating Human Factors Guidelines for Integrating Speech and Manual Input in CAD Systems**	**244**
	Overview	245
	11.1 Introduction	245
	11.2 Derivation and development of guidelines	246
	11.3 Validation of guidelines	250
	11.4 Analysis and discussion	252
	11.5 Summary	261
	Next chapter highlights	261
CHAPTER 12.	**Thesis Outputs and Future Directions: A Summary and Discussion**	**262**
	Overview	263
	12.1 Introduction	263
	12.2 Thesis outputs: summary and discussion	266
	12.3 Problems encountered	272
	12.4 Future directions in user interface development	273
	Final remarks	276
Bibliography		277
Appendices		299

LIST OF FIGURES

FIGURE 1.1	A model of human-computer interaction science support	24
FIGURE 4.1	Overview of a behaviour-based methodology for evaluating CAD systems	74
FIGURE 4.2	VITAS configuration showing the hardware link	82
FIGURE 5.1	A generic production showing the possible conditions and actions	101
FIGURE 5.2	Control structure architecture: knowledge executor	105
FIGURE 5.3	System behaviour blackboard representation at input/output level	108
FIGURE 5.4	Contextualisation of system behaviour blackboard within a superordinate design process blackboard	111
FIGURE 6.1	Visual behaviour types - eye gaze to specific targets (n=14)	121
FIGURE 6.2	Eye transitions between input-output devices (n=14)	123
FIGURE 6.3	Hand manipulation of input devices (n=14)	123
FIGURE 6.4	Right and left hand transitions between input devices (n=14)	125
FIGURE 8.1	Configuration of experimental system	160
FIGURE 8.2	General experimental set up showing a subject using a speech recogniser to perform CAD task	161
FIGURE 8.3	Speech recognition performance of speech input group (n=12)	172

LIST OF TABLES

TABLE 4.1	Example of visual behaviour menu and action/time codes listing	83
TABLE 5.1	A summary of behaviour knowledge sources	109
TABLE 6.1	VITAS menu for scoring behaviour protocol of CAD experts	119
TABLE 6.2	A summary of system characteristics determining subject selection	126
TABLE 6.3	Behaviour knowledge sources of three CAD experts	127
TABLE 6.4	Mean and standard deviations of subjective ratings of system use	130
TABLE 6.5	Utilisation of system information during task performance (n=15)	131
TABLE 7.1	AutoCAD command structure, root menu and major commands	144
TABLE 7.2	A summary of the strengths and limitations of input devices	151
TABLE 8.1	Categories of visual and manual behaviours: Experiment 1	165
TABLE 8.2	Categories of verbal behaviours: Experiment 1	166
TABLE 8.3	Effects of unitary manual and speech input systems on visual behaviour	169
TABLE 8.4	Effects of unitary manual and speech input systems on manual behaviour	171
TABLE 8.5	Effects of unitary manual and speech input systems on performance	174
TABLE 8.6	Correlation of behaviour and performance measures: Experiment 1 results	175
TABLE 8.7	Frequently-used and difficult commands	179
TABLE 9.1a	Categories of manual and verbal behaviours: Experiment 2	196
TABLE 9.1b	Command/data list for verbal behaviour	197
TABLE 9.2	Effects of unitary speech and integrated speech-manual input systems on visual behaviour	200
TABLE 9.3	Effects of unitary speech and integrated speech-manual input systems on manual behaviour	202
TABLE 9.4	Effects of unitary speech and integrated speech-manual input systems on verbal behaviour	204
TABLE 9.5	Verbal content of speech: Experiment 2	204
TABLE 9.6	Effects of unitary speech and integrated speech-manual input systems on performance	206
TABLE 9.7	Correlation of behaviour and performance measures: Experiment 2 results	207
TABLE 10.1	Categories of visual and verbal behaviours: Experiment 3	227
TABLE 10.2	Effects of old and new hybrid systems on visual behaviour	229
TABLE 10.3	Effects of old and new hybrid systems on manual behaviour	232
TABLE 10.4	Effects of old and new hybrid systems on verbal behaviour	234
TABLE 10.5	Effects of old and new hybrid systems on performance	234
TABLE 10.6	Correlation of behaviour and performance measures: Experiment 3 results	236

ACKNOWLEDGEMENTS

I gratefully acknowledge the assistance I received from the following during the course of this PhD.

- Universiti Kebangsaan Malaysia (UKM), for granting me study leave under The Staff Training Programme, and colleagues in the Department of Psychology, UKM, for sharing my workload.

- The Public Services Department, Malaysia, for financial support in the form of a Training Grant.

- The Ergonomics Unit, University College London (UCL), for this opportunity to undertake training in an intellectually stimulating and challenging environment.

- The Department of Psychology, and the Audiovisual Unit, UCL, for all their technical services.

- The Department of Psychology, Birkbeck College, in particular, Dr. Angela B. Summerfield, and Professor J.P. Watson of The United Dental and Medical Schools of Guys and St. Thomas's Hospitals, for the use of VITAS facilities.

- Dr. Jacqueline V. Laws, STC Technology Ltd., for her programming support and invaluable advice on VITAS. Her willingness to give up her time is greatly appreciated.

- The Management of Bechtel, British Petroleum and Fluor Daniel, UK, for allowing access to their design departments. My special thanks to the subjects who participated in the Observational study.

- All subjects who volunteered in the Experiments and the Optimisation study. Their willingness to endeavour patiently in the investigations was a boost to research morale. My heartfelt thanks to each of them, especially Selma Nicholson.

- Professor R. Spence of Imperial College; Mr. P. Windle of Autodesk Ltd.; Mr. M. Talbot of Communication Aids Centre; Mr. G. Ashworth, Mr. P. Rushton, Mr. Tony Lo and Mr. M. Atyeo of the Human Factors Division, British Telecoms, for participating in the Validation study.

- Friends in the Ergonomics Unit, for their encouragement in seeing this thesis to a conclusion. Thanks are due to Miles Benson for his technical help, Andrew Life and Lim Kee Yong for many helpful discussions.

The undertaking of this research enterprise itself was made possible by the guidance and constructive input received from my two supervisors, Dr. Andy Whitefield and Professor John Long. I am greatly indebted to them - Andy for his dedication and exemplary supervision; John for his vision and invaluable advice. Their criticisms and feedback helped to shape the thesis further. Needless to say, any ambiguities that may still remain in the thesis are entirely my responsibility.

I am further grateful to members of my family for their encouragement during the arduous task of producing this thesis. My sister, Azah, deserves special mention for her support when the going was stressful. My greatest debt of gratitude, by far, is to my parents who have been a constant source of inspiration and support in more ways than I could possibly recount, not only during this training, but over many years. Most of all, I am deeply indebted to my husband, Sobri, who suffered three and the half years of enforced separation; his sacrifices and forebearance are indeed an inspiration. It is to them that, with much love, I dedicate this thesis.

CHAPTER 1

Introducing the Research Problem, Solutions and Application Domain

Overview

1.1 Introduction

1.2 Human factors concern

1.3 The problem

1.3.1 Defining optimal behaviour and performance

1.3.2 Evidence of non-optimal behaviour in using manual input devices

1.4 Potential human factors solutions

1.4.1 Using unitary speech input

1.4.2 Integrating speech and manual input

1.4.3 Development of design standards and guidelines

1.4.4. Conclusion

1.5 The application domain

1.6 The research process

1.6.1 Research aims

1.6.2 Research approach

1.7 Organisation of the thesis

1.8 Summary

Next chapter highlights

CHAPTER 1

Introducing the Research Problem, Solutions and Application Domain

OVERVIEW OF THE THESIS

This thesis investigates human factors aspects of integrating the use of speech and manual input devices in human-computer systems. The thesis adopts a research approach that begins with an initial problem specification in the real world, followed by the investigation of solutions in laboratory contexts, then the application of the research findings to develop human factors guidelines and their validation in the workplace. This approach involves identifying and analysing the problem of non-optimal behaviour in using manual only input devices (ie. not combined with other input modes, such as speech), in a way that would permit description of solutions to be directed effectively towards the improvement of behaviour. Potential solutions explored include using unitary speech input and integrating both speech and manual input in a single system. The domain of computer aided design (CAD) is used to exemplify the problem and to explore the potential of each solution toward solving the initial problem.

A blackboard framework of design is used to develop a model of system behaviour. The model is a representation of behavioural elements of a CAD system in the form of knowledge sources, and describes how different knowledge sources are recruited during CAD performance. Variability in system behaviour, in particular non-optimal behaviour, is expressed in relation to the model.

The research findings, derived from three experiments, show that integrating speech and manual input within a single system leads to a reduction in non-optimal behaviour, in addition to improving speech performance. The findings are used to generate some human factors guidelines for integrated computer systems involving the use of speech and manual input. The guidelines are intended for use by end users, CAD implementors and systems designers in configuring and/or designing multimodal CAD systems.

This chapter, in particular, identifies that the unitary use of manual input devices results in non-optimal behaviour and that speech input might offer an alternative solution. It defines input devices and explains their significance in supporting task performance. It also defines non-optimal behaviour using the human factors literature. Human factors solutions to the problems of manual input are described with unitary speech input and speech-manual integration selected for further investigation. The rationale for integrating speech and manual input, and the organisation of the thesis are also presented.

1.1 INTRODUCTION

When people use computers, they are often faced with alternative means of interacting with the system to carry out a given task. A wide range of input and output (I/O) devices are available to support interaction, and the number is continuing to increase. (As used here the term *input* refers to the computer, thus a keyboard, mouse, etc. are types of input devices because they are used to input information to the computer.) The range of input devices, each with its own properties, makes it difficult to determine the appropriateness of a device for a given task or context (eg. Greenstein & Arnaut, 1986; Buxton, 1986; Whitefield, 1986a; Ritchie & Turner, 1975).

The advent of speech recognition technology has provided what may be a more 'natural' mode of control and communication with the computer than the conventional manual input mode such as the keyboard. The particular advantages of using speech to communicate with machines are well documented in the human factors literature (eg. Hapeshi & Jones, 1988; Waterworth & Talbot, 1987; McLeod, 1987; Newell, 1985; McCauley, 1984). Though a nascent technology, the recent use of automatic speech recognition (ASR) in industry (eg. the Ford factory in Germany and General Electric in the USA for quality control and warehouse inspection) and in CAD in particular (eg. cartographers at Clyde Surveys Ltd. and hydrographers at the Ministry of Defence in UK) has helped to save time, in addition to reducing the paper work considerably (Noyes & Frankish, 1987). Research on speech input in applications such as avionics, office-based tasks (eg. text-editing, CAD), industry (eg. parcel sorting) and as an aid to disabled people (eg. Taylor, 1986; Gould & Alfaro, 1984; Martin, 1989; Visick, Johnson & Long, 1984; Damper, 1984), have shown that the use of ASR provides a means of reducing the workload otherwise induced by manual input, thus making better use of the potential of the human's sensory and motor systems. Interest in data entry is no longer confined to manual responses but includes the possibility of combining speech and manual input, integrating both modalities for a single purpose.

Although speech input enables novel and potentially profitable forms of interaction, its use in supporting task performance in any application needs to be assessed. Research into the effectiveness of input devices has a major role to play in the domain of human-computer interaction (HCI), especially in the development and refinement of input devices. The purpose of this development is to create input devices that are ergonomically designed, to suit human physical and cognitive characteristics, and so promote efficient, reliable and even pleasurable input to a system. This philosophy underlies the work to be described here.

A definition of an *input device* might be simply thought of as any instrument, apparatus or mechanism that can be used to enter information into a computer (Booth, 1989, p. 22). (The term *information,* as used in the thesis, includes commands, graphical and alphanumeric data.) In essence, an input device is the physical means by which the user communicates with the computer system. It also transforms information in the interaction process. It follows that if the input device is the communication link between human and machine, its design is crucial for effective task performance.

The importance of input devices in user-computer interfaces is best expressed by Buxton (1986, pp. 321, 336):

> "... when we discuss user interfaces, consideration of the physical transducers too often comes last, or near last. And yet, the physical properties of the system are those with which the user has the first and direct contact... If we are to improve the quality of human-computer interfaces we must begin to approach input from two different views. First, we must look inward to the devices and technologies at the finest grain of their detail. Second, we must look outward from the devices themselves to see how they fit into a more global, or holistic, view of the user interface".

With this view, the next section briefly examines the human factors concept, which includes input devices as one of its four elements.

1.2 HUMAN FACTORS CONCERN

The term *human factors* (HF) has several interpretations, as defined by Meister (1971). In an attempt to provide some consistency and standardization to the term, a committee set up by the Executive Council of The Human Factors Society proposed the following definition:

> "Human factors is that branch of science and technology which includes what is known and theorized about human behavioral and biological characteristics. It serves as a repository and source of data and principles that can be validly applied to specification, design, evaluation, operation, and maintenance of products and systems that are intended for safe, effective, satisfying use by individuals, groups and organisations ... The term*human factors* is considered synonymous with the term *ergonomics*." (Christensen, 1988, p. 9).

This definition highlights the importance of ergonomics data in the design of systems, including information-based systems. Central to human factors is human behaviour, the study of which provides the basis for understanding the problems associated with using a system. Failure to address the human element results in the design of systems that take technology as their starting point (Bjorn-Andersen, 1988; Meister, 1987).

Besides the human element, Meister (1971) described three other elements which influence the efficiency with which people can use equipment to accomplish the functions of their task. Together, the four elements make up a system. The other three elements are:

- **Equipment** (device or machine) - the characteristics of the equipment with which people must interact (eg. the arrangement of controls and displays);
- **Task** - the characteristics of the jobs (or functions) which people must perform in order to accomplish goals;
- **Environment** - the physical surroundings in which the equipment must be operated and maintained (eg. physical layout of the room, noise level, lighting, etc.).

Characterising a system by its constituent parts is consistent with the system definition of Hall and Fagen (1956). They defined a *system* as a set of objects together with relationships between their attributes. The objects are simply the components of a system (eg. computer) and the attributes are the properties of components (eg. behaviours). Relationships tie the system components together, and components are modified through interactions with each other.

Human-computer systems, therefore, comprise the components of human, computer, task and environment. Because system functioning is critically dependent upon the relationships between these components, any characteristic of the device which makes it difficult for users to carry out their task reduces the efficiency of system functioning. Therefore, the human factors practitioner (designers, engineers, etc.) needs to consider the device to ensure that the system under development is designed for most effective use.

In conclusion, input devices have an important role to play in enabling task performance. This means that their design is crucial and research towards refining their development has been one of the many concerns of human factors. The goal of such research is to improve the device's performance (efficiency, utility, etc.) in particular, and the overall system performance in general. This suggests that there is ample scope for research into the effectiveness of input devices.

1.3 THE PROBLEM

This section identifies the central problem addressed by the thesis. As previously mentioned, input devices are an important element in the design of human-computer interface. Every input device specifies some of the requirements associated with its use. With the alphanumeric keyboard, this means the need to learn and acquire some typing or keying skills. Therefore, direct end-users (in Smith's (1980) terms), who do not acquire the necessary skills become 'hunt-and-peck' users (ie. looking at the keyboard for the keys) rather than touch typists (ie. looking at the screen or offline document). Such users are unlikely to be skilled at using the keyboards, let alone skilled enough to be able to type without visual feedback (Long, Nimmo-Smith & Whitefield, 1983). Besides the key-board, the use of the graphics tablet in CAD tasks also requires a division of attention between attending to the visual display and operating the tablet. The notion of 'eyes-and-hands busy' refers to this phenomenon. This term, often cited in the literature (eg. Martin, 1989; Schmandt, 1985; Morrison, Green, Shaw & Payne, 1984; Damper, 1984) is not well defined in terms of the relationship between device use and user I/O requirements. In other words, the extent to which particular input devices require visual attention is not clearly specified.

The frequent visual tracking of key pressing or graphics tablet activity is considered here to be non-optimal behaviour. There are two possible reasons for suggesting that off-screen gazing (ie. looking away from the screen(s) to manipulate input devices) can be a problem. First, the frequent eye transitions between screen and device may divide the attentional resources required for the task. This would incur performance costs, thereby reducing efficiency. Second, frequent changes in the direction of visual orientation can be physically stressing, thus imposing anatomical and biomechanical demands on the user. The problem can be further aggravated if two input devices which share the same modality are used (eg. keyboard and mouse). In addition to increased eye transitions, there will be frequent inter-device movements of the hand(s). Like eye transitions, hand transitions too incur performance costs. In short, frequent movements of the head and/or the hands to meet the requirements of the input devices will produce non-optimal behaviour, resulting in performance costs.

1.3.1 Defining optimal behaviour and performance

The above statement of the problem requires that performance and behaviour are clearly defined. Within HCI, the term *behaviour* as defined by Dowell and Long (1988, p. 22) expresses the means by which the system accomplishes its task. The system performs tasks, achieving task goals within application domains. Computer aided design is concerned with systems whose components are users (designers) and computers (I/O devices). The interaction (ie. the mutual influence of user behaviour and of computer behaviour) determines system behaviour. It follows that *system behaviour* can be ex-pressed in terms of system component behaviours - user behaviour and computer behaviour. (Following the description of human I/O channels, user behaviour is categorised here into visual (eye), manual (hand), verbal (speech) and cognitive (memory) behaviours, excluding auditory.)

The term *performance* "expresses the effectiveness of the system in accomplishing tasks... in terms of the quality of the task product [quality describes the actual product of a task with respect to the desired product]... and the incurred resource cost of production" (Dowell & Long, 1988, p. 21). Therefore, a system performs well if it enables the task to be accomplished with quality assured in the output and with minimal costs of producing it. In addition, a system that performs well is likely to be accepted by the user. Therefore, user acceptability measures the affective costs of system performance (see Dowell & Long, 1988).

However, it should be noted that: (1) performance is different from, but determined by, behaviour; (2) many different behaviours can produce the same performance; and (3) behaviour may be more or less optimal with respect to performance. (In this study, the term performance is used generally to mean system performance, unless specified otherwise as pertaining to either user or device performance per se.)

The dictionary definition of *optimum* is "that point at which any condition is most favourable" (Chambers, 1985). It is a precise word most suitable to apply to a single member of a well-defined set of objects or events with respect to a well-defined criterion (Sheridan, 1988). By this definition, optimal behaviour is the most favourable condition of behaviour with respect to a well-defined criterion. On the basis of criteria, such as the number of times (ie. absolute frequency) and the length of time (ie. relative duration) of a specific behaviour (eg. eye transition to a target), it is possible to demonstrate that the use of some input devices results in non-optimal behaviour. The relationship between such behaviour and performance is explained in Chapter 4. Other terms used in the thesis which require definition are also dealt with in Chapter 4.

The next section presents some evidence of non-optimal behaviour as documented in the literature.

1.3.2 Evidence of non-optimal behaviour in using manual input devices

Several studies in HCI have examined the performance aspects of using input devices in a variety of applications (eg. Karat, McDonald & Anderson, 1986; Card, English & Burr, 1978; Whitfield, Ball & Bird, 1983; Haller, Mutschler & Voss, 1985; Johnson, Long & Visick, 1985). These studies, however, do not address behaviour in the way it is addressed in this thesis. But Card, Moran and Newell (1983) predicted the performance times of using input devices in relation to the GOMS (Goals, Operations, Methods, Selection) model. Through formal task analysis (of a real CAD VLSI circuit task), in which an expert CAD operator's behaviour was broken down into a number of cognitive cycles and motor components, they have shown that "the physical operations alone account for 96% of the user's execution time, which leaves little time for mental operations" (Card et al., 1983, p. 353).

Previous empirical research which examines behavioural aspects of device use is relatively sparse (eg. Rickett, 1987; Van der Heiden & Grandjean, 1984; Bolt, 1980). Van der Heiden and Grandjean (1984), for example, studied 38 CAD operators performing three different tasks, that is, mechanical design, printed circuit board design and electrical schematics. The results showed that CAD operators watch the screen 46-68% of the time, operate the keyboard 14-22% of the time, operate the graphics tablet 26-48% of the time, and manipulate the document 9-15% of the time. Depending on the task, operating the keyboard and graphics tablet accounts for between 40-70% of their total worktime at the terminal, relative to other task-related activities.

Furthermore, "the use of two input media - tablet and keyboard - gives rise to interference problems. Since the tablet was the primary input medium, the keyboard was usually left at the side of the tablet" (Van der Heiden & Grandjean, 1984, p. 344). In the case of keying large quantities of data (ie. alphanumeric entry of text via the keyboard), 41% of the operators placed the keyboard on top of the tablet to avoid the awkward positioning of arms in order to reach it. Monk (1986) claimed that moving a hand from the keyboard to a pointing device (eg. mouse) will disrupt typing performance especially that of touch typists for whom hand position is critical. These inter-device transfer times incur behavioural costs, namely, (1) shifting the hand(s) between two manual input devices, and (2) switching attention from one device to the other.

Besides the anatomical demands that these devices place on the user, the above studies clearly illustrate two points. First, the use of off-screen input devices, particularly keyboard and tablet, requires some degree of visual monitoring, thus keeping the eyes busy. Second, the use of two manual input devices requires the physical movement of at least one hand between the keyboard and tablet in changing from one operation to another, thus keeping the hands busy. Both types of behaviour - off-screen gazing and between-device hand transitions - are regarded here as non-optimal behaviours.

However, it is not clear from Van der Heiden & Grandjean's (1984) study what is meant by 'interference problems'. It is plausible that the interference could refer to: (1) the physical arrangement of input devices in relation to the task; (2) the inter-device hand movements; and/or (3) the division of attention between I/O devices. Also, it is not clear what effect eye and hand transitions have on task performance. In other words, the relationship between user behaviour and task performance as a function of input devices is not sufficiently documented.

In conclusion, the problem identified above concerns the unitary use of manual input devices for performing CAD tasks. It is evident that such use incurs frequent eye and hand transitions, which in turn incur performance costs to the user. The aim of this thesis is to analyse the nature and extent of the problem and to investigate solutions that might alleviate it. The next section considers some possible solutions and identifies those for further investigation.

1.4 POTENTIAL HUMAN FACTORS SOLUTIONS

1.4.1 Using unitary speech input

The first potential solution to be considered involves substituting manual input by speech. The idea of speaking to a computer seems particularly appealing. The growing interest in speech as an interaction medium may stem from a desire to communicate with computers as easily and quickly as communicating with people (Newell, 1985). Speaking is a well developed skill. The cognitive processes involved in speech production are so well automated that one can speak while carrying out other tasks using the hands, eyes and feet (eg. driving).

Speech input in computer systems has the further advantage of freeing users from their keyboards so that the hands can undertake other tasks, such as sketching a design, or leafing through reference material. Some users have never learned to type efficiently, and even if they have, typing is slower than speech (Martin, 1989). Therefore, in multi-task situations, speech provides an additional response channel over which the workload can be spread (eg. Leggett & Williams, 1984; Poock, 1982; Tsang, Hart & Vidulich, 1986).

Despite considerable efforts to adopt speech in a range of applications, its use is still confined to a few, select applications. Much of the reason for this is that the performance levels of speech recognisers have not improved in accordance with predictions (Talbot, 1987a). Almost all of the systems developed to date require trained speakers - people who speak to the machine carefully (Hauptmann & Rudnicky, 1988). Unfortunately, people are not consistent in the way they speak, and current systems are not robust enough to cope with much variation and change in human speech (Waterworth & Talbot, 1987).

There have been few successful attempts to have the speech recognition system adapt to the speaker, and so the onus in overcoming poor recognition rates has been on the user, who has to adapt to the system. This adaptation can be considerable, and in itself can cause the user to lose confidence in the system (Martin & Welch, 1980; Talbot, 1987b). Therefore, it is envisaged that poor speech performance may impose problems for the user in terms of behaviour. In other words, this solution might incur some costs to the user.

As claimed by Hapeshi and Jones (1988, p. 251), "... speech interfaces are far from being 'natural' and 'usable'. Until the careful application of ergonomic principles in their design and implementation, ASR systems are unlikely to compete with proven manual input devices". Therefore, some solutions need to be found for optimizing its use with other input devices.

1.4.2 Integrating speech with manual input

Given the state of the art of speech recognition technology, unitary use of speech input may still be a long way off. As indicated above, researchers (eg. Taylor, 1986; Waterworth & Talbot, 1987) have claimed that speech recognition can have a significant role in computer interfaces, provided that it is properly integrated with the rest of the interface. There are a number of reasons and evidence to support this claim.

First, the use of complementary input devices that are compatible in their rela-tionship with each other is widely acknowledged in the literature as being promising. The selection of the most appropriate sets of input devices in any instance will depend on the nature of the tasks to be performed as well as on the users and the environment. For many applications, the relatively low interference between voice and pointing devices, coupled with the fact that only one requires a manual response, makes the combination a promising one (Whitefield, 1986a). Studies of dual-task performance (eg. Vidulich, 1988; Damos, 1985) have demonstrated improved performance when one task required manual response and the other, speech.

Second, there is evidence (Wickens, Vidulich & Sandry-Garza, 1984) suggesting the existence of a unique compatibility relation between modalities of input (ie. visual) and output (manual, speech) and codes of central processing (spatial versus verbal). For this reason, perhaps, the VDU in interactive systems is used predominantly as an output device while the keyboard and other manual tools are used as input devices.

A third reason is the suggestion that combining speech and manual input can impose a high degree of pyschological organisation, particularly if commands are spoken, and data typed or digitised. It appears that well-organised response sets are easier to learn and allow faster, more accurate performance (Morrison et al., 1984). Besides, a speech-plus-manual configuration might have certain advantages for the 'casual' user: notably it might achieve a classic separation of function by modality.

Lastly, the use of a 'modeless' user interface can help to resolve the ambiguity which gives rise to mode errors (Monk, 1986; Thimbleby, 1982; Poller & Garter, 1984) or 'slips of action' (Norman, 1983). This can be achieved by: (1) increasing the bandwidth of the user interface, that is, introducing new keys or additional input devices as alternatives to the keyboard; or by (2) rationalizing the functionality of the system. An example of the latter is the keyboard may be reserved for text entry and an alternative input device, say speech, for command specification. A rationalized system may offer fewer alternatives but it is 'simpler', incurring lower cognitive overhead (ie. keeping track of the mode one is in), and is therefore easier to learn and easier to use (Monk, 1986).

Given the above, there is reason to believe that integrating speech and manual input might prove a better solution than using speech input on its own.

1.4.3 Development of design standards and guidelines

Design standards, by definition, imply the minimum acceptable requirement for design, whilst *guidelines* are recommendations or rules of thumb for guiding the design process. Meister (1985, p.6) defined design standards as consisting of three elements:

(1) A relationship between a system characteristic and human performance, expressed ideally in the form of a task-accomplishment metric;

(2) Data describing what the effect of a non-optimal relationship would be, again in terms of task failure or error consequences; and

(3) Some principles for incorporating a desired system characteristic into design.

The need for design standards, such as those by the American National Standards Institute (ANSI) and International Standards Organisation (ISO), is much emphasised in the literature, particularly standards relating to input devices. In the absence of design standards, the designer is forced to make choices involving difficult tradeoffs, for example, between hardware features and software support facilities. Also, lack of standards inhibits the sharing of equipment and programs and results in much duplication of effort.

In general, guidelines relating to the design of physical interfaces are more readily available than those concerning cognitive interfaces (Nickerson, 1986). The literature provides a rich source of guidelines, for example, in (a) dialogue design (eg. Shneiderman, 1980; Williges, Schurick, Spine & Hakkinen, 1986; Cole, Lansdale & Christie, 1987) and (b) computer graphics (eg. Newman & Sproull, 1979; Davis & Swezey, 1983). But guidelines for determining the relative merits of the different devices for specified types of application (with the exception of the keyboard) have yet to emerge. There is general agreement that most of the design guidelines that have been suggested, however, lack empirical support (eg. Meister 1988; Nickerson, 1986).

In short, development of design standards is one possible solution to some HF design problems, but would be a difficult one, particularly the determination of the consequences for performance of a non-optimal design. However, by experimentation it is possible to evaluate the relationship between a system characteristic and resultant user behaviour and performance. The outcome of such investigation could be applied to develop some human factors guidelines. This solution will be considered in the thesis.

1.4.4 Conclusion

Three possible solutions to the problem of unitary use of manual input devices have been suggested. The first and second solutions, that is, using unitary speech input and integrating speech and manual input directly address potential ways of overcoming the problem. The third solution, that is, the development of guidelines, is an application of the findings from the investigation of the two solutions. These guidelines might be used to develop new systems which integrate speech and manual input.

Other potential solutions, but beyond the resource capabilities of this thesis, include the design of interactive systems that is driven by natural language. The latter is among the possibilities for widening the I/O bandwidth of the interface.

1.5 THE APPLICATION DOMAIN

This section considers briefly the choice of computer aided design (CAD) as the context for the integration of speech and manual input.

CAD is a technology to support the application of computers to design. Increasing design productivity is clearly an important potential outcome of CAD systems. Comparing the output between conventional drawing board and CAD, it has been reported that CAD can increase productivity from between 3:1 to more than 10:1 ratio (Groover & Zimmers, 1984). This perhaps accounts for the increasing use of CAD in a wide range of industrial settings (see Chapter 3). Current CAD systems typically employ input devices such as the alphanumeric keyboard, graphics tablet (also known as digitiser), mouse and joystick for the control of display elements. Of these, the tablet-plus-keyboard configuration has been the most common.

It was said previously that the use of two or more manual input devices may result in non-optimal behaviour, and it has been suggested that speech applications would suit tasks that: (1) require considerable involvement of the users's hands and eyes; (2) involve a limited set of discrete operations; and (3) are not highly paced (Clark, 1986; Welch, 1980). CAD also requires the input of different types of data, and combinations of input devices have been found to be most appropriate for such data entry tasks (see Chapter 3). Therefore, on these criteria, CAD appears to be a good candidate for speech application. Most previous work on speech input in CAD (eg. Kelway, 1986; Green et al., 1983) has focused on hardware developments rather than on usability aspects of speech use. Thus, there is much scope for research.

In conclusion, speech input has potential in CAD applications as a means of freeing the eyes and hands to perform other concurrent tasks. Because of current technological limitations, speech input is only rarely encountered in CAD applications and is not yet a serious alternative to the tablet and keyboard. If speech input can replace keyed and/or digitised input, then the problem of non-optimal behaviour as discussed earlier might be reduced.

1.6 THE RESEARCH PROCESS

1.6.1 Research aims

The thesis aims to address the issue of integrating speech and manual input devices in human-computer systems. To achieve this aim, a methodology for evaluating systems in the CAD task domain is required. The methodology involves:

(1) understanding better the capabilities and limitations of the human in relation to the demands of the computer, in particular the input devices;

(2) developing data analysis techniques, including:

  (a) a modelling method which could serve as an explanatory tool for non-optimal behaviour,

  (b) a computerised technique for analysing behaviour protocols which would provide the data for statistical analysis and for the modelling procedure, and

  (c) a statistical tool for testing the effect of types of systems on user behaviour and/or performance.

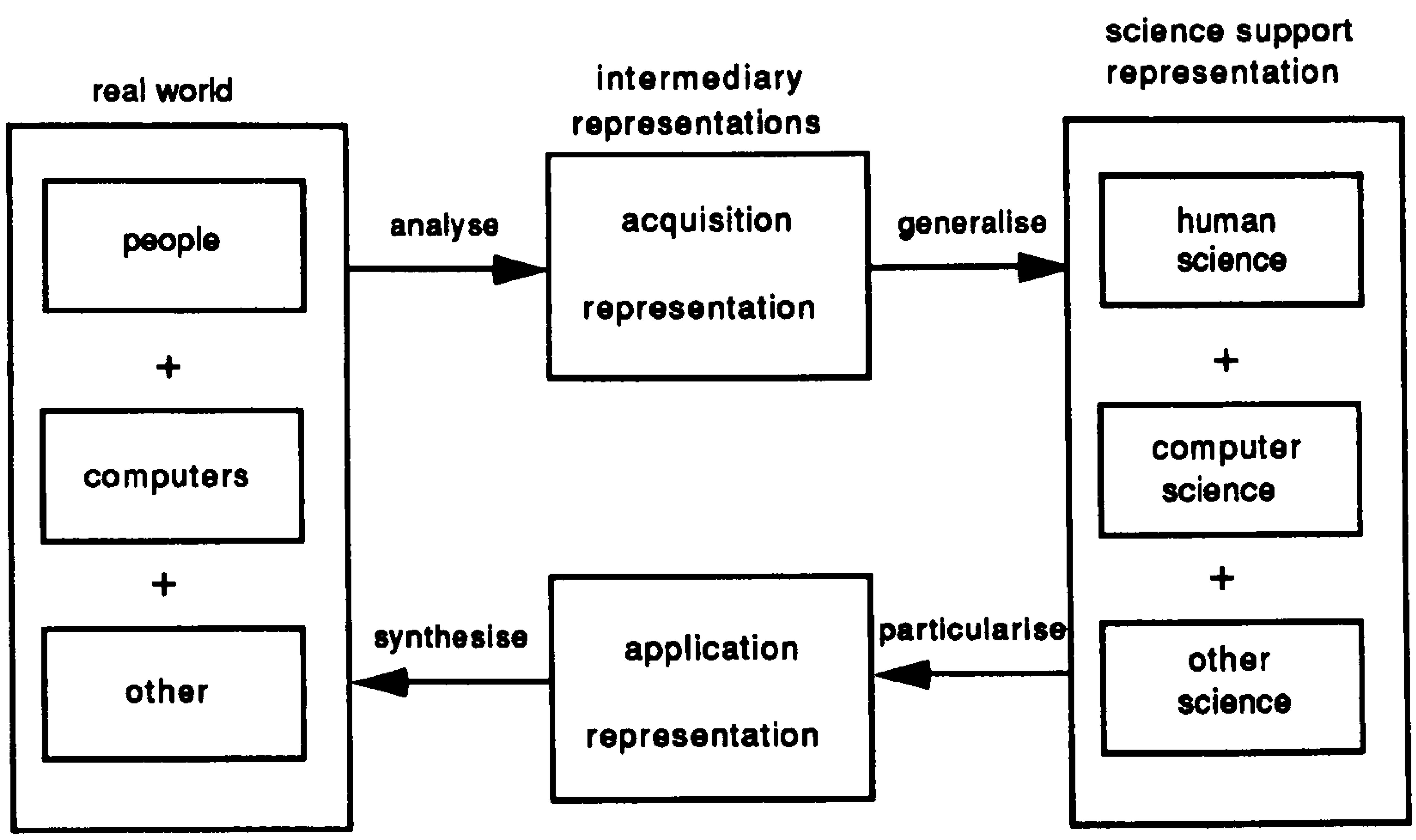
Further description of the methodology is given in Chapter 4.

1.6.2 Research approach

In terms of the framework for ergonomic activities developed by Long (1987a; 1989), this research comprises four types of activities: (1) analysis; (2) generalisation; (3) particularisation; and (4) synthesis. The order in which the activities occur is not necessarily sequential, but overlaps.

With reference to Figure 1.1, the activity of *analysing* the real world (ie. observation of CAD experts at work) produces an *acquisition representation* which supports a laboratory simulation for experimentation (ie. specification of problems and investigation of solutions). The activity of *generalising* the experimental findings produces the *science support representation* (ie. general principles of device integration and system development). The activity of *particularising* the science representation produces an *application representation* in the form of a blackboard model of system behaviour and some human factors guidelines for configuring or designing integrated speech-manual input systems. Finally, the activity of *synthesising* this activity with the real world (ie. validation of guidelines in the workplace) contributes to changed computerised systems, that is, interaction development practice, as suggested by Long (1989).

In applying this framework, the thesis represents a complete cycle of research development from problem specification in the real world to synthesis in the workplace. Within the resource constraints of the research, it is not possible to address all activities in depth. This means each activity is addressed selectively, with particular emphasis on the analysis



Source; Long, J.B. (1989). Cognitive ergonomics and human-computer interaction: an introduction. In J. Long and A. Whitefield (Eds.), *Cognitive Ergonomics and Human-Computer Interaction*. Cambridge: Cambridge University Press.

Figure 1.1. A model of Human-Computer Interaction Science Support

and generalisation. In short, an attempt is made to address a subset of the total spectrum of problems, solutions and applications in greater detail.

1.7 ORGANIZATION OF THE THESIS

The thesis is divided into 12 chapters. The overview at the outset of each chapter gives a summary of the chapter content. The summary at the end of each chapter presents the principal discussion points and findings which will relate the problem to the solutions.

This chapter sets the scope of the thesis and identifies the problem and solutions for investigation. It also defines input devices and their connection with non-optimal behaviour. The requirements of a methodology for evaluating CAD systems are specified and the general research approach is described. Chapter 2 presents some human factors issues of speech and manual input devices. Previous studies which have examined these issues are reviewed. In Chapter 3, development in CAD technology is reviewed, including its history and impact in advanced and developing countries. Since CAD is part of a design process, a brief introduction to design is relevant and this is also made in Chapter 3.

A behaviour-based methodology is described in Chapter 4. To provide clarity, terms used in the thesis are operationally defined in the same chapter. Part of the methodology is the development of a modelling technique for knowledge description. This is described in Chapter 5 in which a framework for modelling system behaviour, using the blackboard model of design, is presented. Data for constructing the 'base' model are provided by an observational study of CAD experts at work. Chapter 6 describes this study and the problems documented on the unitary use of manual input devices. To investigate the problem further, a 'demonstrator' CAD system is required. Specifications for this system are outlined in Chapter 7, including a study conducted to optimise the experimental system to be used in the research.

Chapters 8, 9 and 10 present the findings of three experiments, aimed at evaluating the potential of three different solutions to the problems documented. The findings from these experiments are applied to develop some guidelines which are described in Chapter 11. A summary and discussion of the thesis outputs is presented in the final chapter, 12. This chapter also reviews problems arising from the work and the limitations of the research approach. The review will also consider the extent to which the aims of the thesis have been accomplished and makes suggestions for further research.

1.8 SUMMARY

This chapter identifies the problem to be addressed in the thesis and selects a particular example of the problem for investigation. Hence, the problem of using manual input devices on a unitary basis within CAD systems is identified. This problem concerns the non-optimality of behaviour in manipulating the input devices to perform CAD tasks, resulting in frequent off-screen eye transitions and between-device hand transitions. The thesis suggests speech input may be a solution and starts to document the problems with its use. Like manual input, speech input has problems which mean it may not on its own be an acceptable solution. This then suggests integrating speech and manual input into a single system.

NEXT CHAPTER HIGHLIGHTS

In Chapter 2, ergonomic issues concerning the use of speech and manual input devices are reviewed. The aim is to provide a characterisation of these devices based on their functionality and applications.

CHAPTER 2

A Review of Human Factors Issues of Speech and Manual Input

Overview

2.1 Introduction

2.2 Background to speech recognition

2.3 Some ergonomic aspects of ASR systems

    2.3.1 Speaker dependence

    2.3.2 Speaking mode

    2.3.3 Enrolment

    2.3.4 Speech performance

2.4 Sources of error in ASR use

    2.4.1 User characteristics

    2.4.2 Task characteristics

    2.4.3 Device characteristics

    2.4.4 Environmental characteristics

2.5 Feedback and backup facilities

2.6 Background to manual input

2.7 Some ergonomic aspects of manual input devices

    2.7.1 Accuracy and resolution

    2.7.2 Feedback mechanism

    2.7.3 Motor movements

    2.7.4 Learnability

    2.7.5 Conclusion

2.8 Review of studies on speech and manual input

    2.8.1 CAD applications

    2.8.2 Non-CAD applications

    2.8.3 Conclusion

2.9 Summary

Next chapter highlights

CHAPTER 2

A Review of Human Factors Issues of Speech and Manual Input Devices

OVERVIEW

The growth of interactive computing has been accompanied by the development of numerous input and output devices through which the user can interact with a computer. In this chapter, human factors issues concerning the use of input devices are reviewed. In Chapter 1, it was said that unitary use of manual input devices can result in non-optimal behaviour. A possible solution to this problem is to use unitary speech input. The review, therefore, examines the utility of speech and manual input in sufficient detail in order for the important aspects of the technology to be appreciated. The chapter concludes with a review of selected studies that examine these issues.

2.1 INTRODUCTION

As mentioned previously, users communicate with computers via physical media termed *input* and *output devices* (I/O). The input device allows the user to enter information relevant to the task into the computer; the output devices (a) display information, either on a video display or via a hardcopy medium, and (b) convey information via speech output. Any weaknesses in this communication link reduce the effective utilisation of these devices by the user.

Often, a specific computer is used by a number of people for a number of tasks, each with their own demands. One approach to dealing with the diversity of demands is to supply a number of input devices, one optimised for each type of transaction. However, as the number of devices increases, the benefits of the approach would generally break down. A more realistic solution is to get as much generality as possible from a smaller number of devices (Buxton, 1986). Devices, then, are chosen for their range of applicability. One example is the graphics tablet which can emulate the behaviour of a mouse. But unlike the mouse, tablets can also be used for digitising predrawn diagrams.

This review of input devices is in three parts. The first part of the chapter examines speech recognition technology, the second part presents manual input technology, while the remaining part considers some studies related to these input devices. The review of speech technology focuses on speech input (ie. speech recognition), while speech output (ie. speech synthesis) will only be mentioned in general.

2.2 BACKGROUND TO SPEECH RECOGNITION

The term *speech recognition* is sometimes used interchangeably with the term *voice recognition*. This thesis will use the term speech recognition exclusively.

Speech is an attractive channel for a range of communication tasks with computers (Lea, 1980; Martin, 1976). The numerous reviews in the literature (eg. Lea, 1980; McCauley, 1984; Simpson et al., 1985; Vaissiere, 1985; Clark, 1986; Starr, 1987; Noyes & Frankish, 1987; Waterworth & Talbot, 1987; Hapeshi & Jones, 1988), covering different aspects of speech I/O technology, from hardware development to dialogue design and applications, indicate the growing interest in speech I/O interface. But merely adding speech to an application does not make it a 'conversational' interface. Voice interaction may be conversational and forgiving, or terse, staccato commands and replies (Schmandt, 1985). That the latter in fact is far more common in computer systems is a reflection of the capabilities of current technology, and in part explains the very limited number of speech systems in actual use. A number of the capabilities that are commonly ascribed to speech technology have been shown to be fallacious, or in general to be overestimated (Talbot, 1987a).

Many of the predictions that were made about the potential uses of speech technology concerned its application in the electronic office. The use of ASR for recording and entering data to a computerised system is ideal in situations where the operator's hands are full, and/or where their line of sight either must be directed away from the input device or is in some way impaired (Visick et al., 1984). Such situations seldom exist in office-based tasks, hence there are few successful applications of speech input in the office, as compared to other settings, such as manufacturing and inspection.

Interest in ASR started some 35 years ago, mainly with the availability of electronic hardware to perform spectrum analysis of signals. Early research on speech recognition was primarily motivated by two different purposes: the automatic transcription of the incoming speech signals into phoneme-like symbols (eg. phonetic-typewriters), or the direct identification of words to command machines by voice (eg. isolated word speech recognisers). With the growth in the use of digital computers in the early 1960s, the prospect of using speech as an input to a computer led to renewed interest in the speech field (see Vaissiere, 1985; McCauley, 1984) and the development of low-cost recognisers that could accommodate a limited vocabulary size.

In 1971, speech recognition research was given a significant boost when the Advanced Research Projects Agency (ARPA) in the USA sponsored the Speech Understanding Research (SUR) project. The HEARSAY and HARPY systems (see Lea, 1980; Barr & Feigenbaum, 1983) are products of this initiative. The late 1970s was characterised by advancements in connected speech recognition, and in 1978, the first connected speech recogniser became available commercially. To date, efforts are continuing in the advancement of phonetic speech recognisers and the design of natural language dialogue for conversational interfaces (Talbot, 1987a).

2.3 SOME ERGONOMIC ASPECTS OF ASR SYSTEMS

Speech recognition systems vary with respect to several parameters (see McCauley, 1984; Simpson et al., 1985; Hapeshi & Jones, 1988). This review is not intended to consider all ergonomic aspects, but only those that are important to the thesis. The order in which the aspects are presented is necessarily arbitrary because each relates to the other.

2.3.1 Speaker dependence

Speaker dependence refers to the extent to which the system must have data about the voice characteristics of the particular human speaker(s) using it. Thus, a speaker dependent recogniser is more particular, able to understand only the individual who trained it. In general, speaker independence is much harder to achieve, and these recognisers exhibit smaller vocabulary and/or poorer recognition.

2.3.2 Speaking mode

Speaker mode refers to the manner in which utterances are spoken to the system. *Isolated* or discrete word systems require each word to be spoken separately, with a slight pause (not less than 100 msec) inserted between vocabulary items. *Connected* word systems are able to recognise longer sentences or groups of words spoken naturally, and identify each individually. The term *continuous speech recognition* actually refers to connected word recognition, particularly for recognition of utterances spoken with natural speech rhythm and intonation (prosodics).

2.3.3 Enrolment

Enrolment or training is the process of providing templates (ie. specific sample utterances) to the recognition system for the different vocabulary items. Speaker-dependent recognisers must be trained separately for each speaker who will use them. In addition to template training, the user is also taught how to get the best recognition performance (termed *user training*) and how to use the system in general (termed *system training*). Most systems provide a procedure for both forms of enrolment. In operational terms, the three types of training are interrelated because they occur together when the user is initially exposed to the system.

2.3.4 Speech performance

Manufacturers of ASR systems usually describe the performance of their systems in terms of the percentage of utterances which are correctly recognised. Average recognition rates are usually quoted by manufacturers as being in the region of 95-99%, and there is very little variation in this figure amongst devices (Talbot, 1987a). Even the more recently developed continuous speech recognisers that are capable of recognising short phrases of speech up to two seconds in duration claim performance in the region of 95-97% (Starr, 1987). Because recognition tests are usually conducted under pristine conditions, "it is misleading for potential users... to believe that they will achieve similar accuracy in an operational situation" McCauley (1984, p. 149).

Other factors that could affect the test include: (1) the nature of the words in the vocabulary: acoustically similar words will cause more substitution errors (Green & Clark, 1981); (2) the speakers used in the tests: there is considerable between-speaker variation in recognition figures due to the sex of the speaker and other demographic variables (Waterworth, 1984); (3) the procedure for creating the voice templates: the number of times each word is uttered, the order in which the words are repeated and the pace of the template creation procedure (Martin & Welch, 1980).

Given the above, Talbot (1987a) pointed out that clearly there is a mismatch between claimed and actual speech performance. Hence, it should be noted that speech performance figures are likely to differ: (1) between manufacturer's claims and device training; and (2) between device training and usage.

2.4 SOURCES OF ERROR IN ASR USE

Recogniser errors may be of three types (Williamson & Curry, 1984): (1) *substitution* (or misrecognition) errors, when the ASR device 'recognises' the wrong vocabulary word; (2) *rejection* (or nonrecognition) errors, when properly spoken words that are part of the active vocabulary are rejected; and (3) *insertion* (or spurious) errors, when the device fails to reject non-speech sounds such as breath noises, sighs, etc. by falsely matching them against a stored template.

Lea (1982) identified more than 80 variables that may influence recogniser performance. These factors are categorised here into user, task, device and environment characteristics. Others (eg. Spine, Williges & Maynard, 1984) have different ways of classifying the variables but the human factors involved are essentially the same.

2.4.1 User characteristics

The most common source of recognition errors is the result of an inconsistency between verbalizing items during training and verbalizing them during operational use of the system by the user (Waterworth & Talbot, 1987). This distinction is crucial in evaluating the performance of ASR systems. In general, user inconsistency is due to:

Stress and fatigue

Users tend to suffer from vocal fatigue, but it is also possible that there will be a general psychological fatigue, and in both cases these would affect the 'task' voice. Related to this are memory failures associated with the size of the vocabulary. The need to remember the vocabulary necessary to use the system can to some extent be alleviated by allowing users to use terms with which they are familiar.

Attitude and expectation

As with much novel technology, users tend to have unrealistically high expectations of the system. Also, the time it takes for users to adopt a more appropriate and consistent attitude towards a system varies between users. It is important that the user is aware of how the system works and that they should not try to 'help' the system. Nye (1982) pointed out that some of the most common errors which users made include over-pronounciation of utterances in order to 'help' the device. To avoid errors caused by behavioural inconsistencies between template training and later use, adequate feedback should be given.

Experience

Experience with ASR systems is known to have a major influence on the recognition accuracy rates, with experienced users being more consistent than relatively naive users (eg. Poock et al., 1982). A distinction can be made between: (1) those experienced in the use of computer plus speech systems; (2) those experienced in the use of computers but not speech systems; and (3) those completely inexperienced (ie. naive). Users in the first category are able to maintain the required uniformity of pronunciation during training and task execution, and achieve better recognition rates than naive users (Talbot, 1987b). Users in the second category may well adapt to speech systems more readily than naive users, but may expect to use the whole computer system with speech, and get frustrated when this is not possible (Harrison et al., 1986). An inexperienced computer and/or speech system user may be nervous or anxious and might assume that the computer can 'understand' more than it does (Hapeshi & Jones, 1988).

In many cases failure to use ASR systems consistently is caused by demographic and physical characteristics such as age, gender, regional accent and speech defects (Lea, 1982). Usually, ASR devices are developed around the voices of the system engineers, typically male, adult voices with relatively standard accent (eg. Hapeshi & Jones, 1988), and hence are not representative of the user population.

2.4.2 Task characteristics

Task characteristics relate to vocabulary size and task knowledge.

Vocabulary size

The number of substitution errors can be reduced by keeping the number of words in the vocabulary to the minimum required by the task, and also by ensuring that the chosen words are acoustically dissimilar (Talbot, 1987b). The number of words that can be handled by commercially available speech recognisers is typically between 50 and 150 at any one time. However, the overall capacity of a recogniser can be much greater than this (Talbot, 1986). With a larger vocabulary more confusions are likely (Lea, 1982) and increasing the number of templates per utterance increases the chances of correct recognition (Spine et al., 1984). Usually 3-10 samples of each word are required. The requisite closeness of the match, the 'reject threshold criterion', can be preset by the user to suit the characteristics of the task.

Task knowledge

There are two aspects of task knowledge. The first concerns training the user on how to use the device so as to maintain a consistent 'task' voice during performance. This not only provides a knowledge of what constitutes successful recognition but also the procedural knowledge involved in device use. The second relates to the actual task *per se*. By clearly specifying what the task entails and what is to be expected, users may be motivated to concentrate on the task itself and not the device.

2.4.3 Device characteristics

Device characteristics relate to template training and the positioning of the microphone that is used in ASR devices.

Template Training

Virtually all ASR systems are based on the principle of template-matching (ie. acoustic pattern matching). With many ASR systems, template training is carried out under very different circumstances to those in which the system will be used (Green et al, 1983). For example, template training might be carried out under relatively quiet conditions, while the system would be used in noisy, busy environments. Another source is that training may be an artificial and repetitive process, with speakers often required to repeat items from a list. Instead, during usage the user may be involved in a more natural, continuous dialogue which may influence the way the user pronounces the vocabulary items.

Microphone placement

Variations in microphone placement can result in differences in the acoustic input to the system

(McCauley, 1984). The most common instruction for a headset microphone is to place it near the speaker's lower lip, slightly off to one side. Once the position is attained, it should be maintained consistently throughout the use of the device.

2.4.4 Environmental characteristics

Environmental characteristics relate mainly to background noise. In avionics, for example, noise can be characterised by: (1) high levels of acoustic noise; (2) high vibration; and poor voice-communications. On the factory floor and office, noise can be characterised by continuous noise from machinery, air conditioning, or intermittent noises from other people, telephones, printers, etc. Intermittent noise in the office is likely to be difficult to predict because of its variability. A hardware solution to background noise is to use uni-directional noise-cancelling microphones.

2.5 Feedback and backup facilities

Given the various potential sources of errors, appropriate feedback (ie. any information provided by the device which the user can utilise to determine whether an utterance has been recognised correctly) will help to reduce these errors. There is general agreement as to the importance of feedback, but there is little consensus as to the form it should take in order to avoid interfering with the user's primary task (Simpson et al., 1985). McCauley (1984) suggests the use of visually-presented feedback because the information is always available to the user and perceiving it does not interfere with the task. In some applications which have a rigid syntax or a small vocabulary, auditory tones as feedback may be sufficient (Lynch, 1984). Other solutions include combining tones with visual feedback (eg. Gould et al., 1983). For details on different error correction procedures in ASR use, see Hapeshi & Jones (1988), McCauley (1984) and Simpson et al. (1985).

The need for backup devices in ASR systems was also raised (eg. Hapeshi & Jones, 1988; Talbot, 1987a). The term *backup* refers to any input devices provided as an alternative to the speech input. The major reason for providing backup facilities is to support the use of speech input given its unreliable performance. As in the case of feedback, the backup device should be one that is compatible with speech input, and users should be allowed to alternate freely between the two modes.

In conclusion, speech recognition technology is still in its infancy, and is much less advanced than speech synthesis technology. The majority of recognisers are isolated-word speaker-dependent devices that extract features from the incoming speech and match these against prototype utterances collected from the same speaker during a training session. These recognisers are sensitive to various user, task, device and environmental characteristics described above. Due to this, speech performance differs between that claimed by the manufacturers and that obtained during actual use. Therefore, current speech use requires some form of backup device that users could fall back to whenever the system fails to produce good recognition. This facility will be considered in the thesis (see Chapter 10).

2.6 BACKGROUND TO MANUAL INPUT

Advances in manual input technology have made possible some alternative data entry methods to conventional keying, such as pointing and the use of digitising tablets. Almost all manual input devices can be called pointing or 'gesture' devices. For example, the keyboard is in a sense a pointing device; it requires 'pointing' to the alphanumeric symbol to be displayed on screen (Comerford, 1984). To distinguish between types of pointing devices, they can be categorised into: (1) those that involve substantial arm movement in order to cover the full range of locations to be pointed at (eg. mouse, graphics tablet, lightpen); and (2) those that do not require arm movement and can be operated by hand movements only across the full range of locations (eg. joystick, trackball, keyboard).

Pointing devices can also be considered as being direct or indirect (Whitefield, 1986a; Maguire, 1988). When using direct devices, the physical movement takes place towards the actual location of the target in space (eg. lightpen, touch screen), while with indirect devices, the physical movement takes place towards an area which occupies a different location in space, but is mapped on to the target area, from which the user receives visual feedback (eg. mouse, touchpad). The graphics tablet, however, can be used both as an indirect device, to select from or input to the screen, and as a direct device, when used in conjunction with a menu overlay.

From a development perspective, the technology of manual input is fairly stable. Introductions to the technological and behavioural aspects of manual input devices are given in the literature (eg. Milner, 1988; Whitefield, 1986a; Comerford, 1984; Newman & Sproull, 1979; Ritchie & Turner, 1975). There is also a growing literature on the hardware and software aspects of these devices in popular magazines (eg. PC User, PC Week, CAD User) which provide users and designers with practical recommendations of device use. Although new devices are being developed, major advances in the field are unlikely.

Most recent developments are geared toward improving the usability aspects (eg. safety, efficiency, user satisfaction) of the devices in relation to known ergonomic standards. For example, the keyboard is one of the oldest forms of input device, and for many tasks it is still the most efficient. The traditional QWERTY keyboard layout was devised in 1866, and became the most widely used format for typewriters and subsequently for computer keyboards.

Because the QWERTY layout is not the easiest to learn nor the most efficient to use, alternative layouts such as the Dvorak and Alphabet keyboard were developed. Using the Dvorak keyboard, the workload distribution between the hands appears to be more evenly balanced (eg. Kroemer, 1972; Bailey, 1982; Maguire, 1988).

The next section examines some human factors in common manual input devices, in particular, graphics tablet, keyboard, mouse, joystick, lightpen, touchscreen and trackball. The aim is to provide an understanding of the criteria which determine the selection of particular devices used in this research. For discussion of other ergonomic issues not addressed here, see Milner (1988) and Whitefield (1986a). Milner, in particular, discusses the human factors of input devices in terms of four components: cognitive, perceptual, motor and sensory.

2.7 SOME ERGONOMIC ASPECTS OF MANUAL INPUT

Input devices perform three distinct functions:

(1) inputting new information (or *information input* function), such as digitising pre-drawn maps or diagrams, entering freehand sketches, entering system commands via function keys, keying-in text or numerical data, and line drawing;

(2) selecting an item among displayed information (or *selection* function), such as pointing to items on a display, choosing between displayed menu or command options, indicating text, lines, etc. to be altered or erased; and

(3) positioning and moving items on screen (or *positioning* and *moving* functions), such as placing and moving items on screen (eg. dragging, stretching) and tracking symbols on a display.

The first function is sometimes termed 'positioning' and the second function as 'pointing'. Clearly, functions (1) and (3) above are not the same, while (2) is often used interchangeably with the term pointing. To avoid confusion, the above classification will be used in the thesis.

The input devices and their functions will be discussed below in relation to ergonomic aspects - accuracy, resolution, feedback, motor movement and learnability.

2.7.1 Accuracy and resolution

Accuracy refers to the extent to which the information transmitted by the device to the computer is a true indication of the actual position being pointed at. *Resolution* refers to the number of points between which the device can repeatedly discriminate (Whitefield, 1986a, p. 98).

In digitising maps or predrawn diagrams, accuracy is important and the graphics or digitiser tablet is best suited to this. But in selecting drawing entities, the resolution of the tablet is more important than its accuracy. The mouse is a low-accuracy device. As the mouse moves across a desktop there is often slippage, resulting in many different mouse movements required to achieve the same pointer movement on screen. The resolution of the mouse in practice is usually limited by the resolution of its associated raster scan display. The joystick is not a very accurate input device and has low resolution which makes it most suited to coarse pointing or tracking tasks. The lightpen resolution is determined by the screen display. Because of the pen aperture and the distance the pen is from the screen surface, there is a lack of precision. Touch screens (or touch panels) are similar to graphics tablet except that they are sensitive to the touch of a finger, and usually have poorer resolution. Hence, they not suitable for digitising line work. The trackball (or tracker ball) does not have such good control as a mouse. Thus, its resolution is its most important limitation.

2.7.2 Feedback mechanism

There are three main ways that a key on the keyboard can provide feedback to the user without having to look at the screen, that is, pressure (force), displacement and sound. Pressure feedback can vary according to a predefined relationship between the amount of key pressure and key displacement, while auditory feedback can exist in the form of a mechanical click or an electronic sound.

With the graphics tablet, as the transducer (stylus or puck) is moved across the tablet, crosshairs on the screen follow it and this provides visual feedback. This visual feedback mechanism is an essential component of the interaction and has the virtue of eliminating the effect of non-linearities in the tablet. Similarly, the feedback of mouse manipulation is provided on the screen. When the mouse is moved across the horizontal desktop, the pointer moves across the screen, thus providing visual feedback to the user. The mouse itself provides feedback on the activation of the button(s) by the snap action of the button.

Although feedback is direct and fast with lightpens, especially with single selection type tasks, there is the problem of parallax if the user sits at a less than optimal orientation to the display screen. Also, the user's hand and the lightpen will at times obscure parts of the display. Unlike the majority of input devices, the touch sensitive screen has no moving parts, so no tactile feedback can be provided. In the normal mode of operation, touch screens provide visual feedback by highlighting the light button or menu option selected. Whilst this may be adequate as feedback, the heavy use of a system may result in much slower response times. The trackball, however, has no proprioceptive feedback because the ball is fixed into the working surface or alternatively, a container.

2.7.3 Motor movements

Pointing responses involve both gross and fine motor movements (see Chapter 4) for bringing about an input, such as touching the screen with a finger or depressing a key on a keyboard. The keys in a QWERTY keyboard are spaced evenly so that a user can reach all the keys easily by finger movements (as opposed to wrist movements). With the graphics tablet, mouse and joystick, the separation of the display and the surface to be pointed at allows for a more optimal body position. Also, the weight of the device and the part of the arm is supported by the desk. Thus, the risk of fatigue with extended use is greatly reduced. Manipulation of the puck and mouse requires holding the device-body between the thumb and middle finger, leaving the index finger to operate the button.

Lightpens are not natural for freehand input because the operator has to hold the pen in an elevated position at the CRT and this can be tiring after an extended period of time. Touch screens too induce physical discomfort and fatigue caused by regularly reaching out and touching the screen. The regular action of pointing causes local fatigue in both the muscles of the active arm and the shoulders and neck.

2.7.4 Learnability

Clearly, the use of keyboards requires some skills. To reach a level of proficiency requires considerable practice and learning. Unlike the keyboard, the rest of the pointing devices are claimed to be easy to learn. For example, the lightpen and the tablet-stylus draw on existing stereotypes in that all users have used pens and pencils, while the touch screen involves touching the screen. However, this does not mean that training is unnecessary.

2.7.5 Conclusion

Because no single input device is well suited to all of the functions mentioned above, due largely to the different characteristic behaviours of each input device, most computer systems employ more than one device. Generally, the mouse, lightpen and joystick are more suited to the selection and positioning/moving functions, since they can transmit only a single class of information, that concerning location. For the inputting of new information, the keyboard seems well suited, at least for alphanumeric information, while the tablet is suited for entering graphics information as well as for selection and positioning/moving.

With respect to tasks, the accuracy and resolution of the device will at least in part determine its usefulness. Also, proprioceptive perception is required to facilitate speed and/or accuracy (eg. touch typing on a keyboard). Most input devices provide this feedback, with the exception of the trackball. In addition to the ease of use and learning criteria, some devices induce fatigue and physical discomfort when used for extended periods of time. Based on these considerations, different applications employ different configurations of input devices. Common configurations used in CAD are discussed in Chapter 3.

2.8 REVIEW OF STUDIES ON SPEECH AND MANUAL INPUT

This thesis is about the human factors of integrating speech and manual input, with CAD as an application. Therefore, to be included in this review, studies had to meet the following criteria:

(1) the focus should be on the use of speech and manual input devices (keyboard, graphics tablet and/or mouse), in office-based tasks, and specifically in CAD;

(2) the data are gathered through experimentation and the methodology clearly reported; and

(3) the concerns are with human factors aspects of device use.

The purpose of these criteria is to delimit the review to research that is relevant and has empirical support. This, then, excludes studies that are concerned with: (a) speech recognition development *per se*, such as hardware (eg. Green et al., 1983; Kelway, 1986; Spine et al., 1984), user characteristics in terms of speaker style (eg. Baber & Stammers, 1989), or interaction dialogue (eg. Bolt, 1980; Schmandt & Hulteen, 1982); and (b) manual input devices *per se* (eg. Karat, McDonald & Anderson, 1986; Card, English & Burr, 1978; Whitfield, Ball & Bird, 1983).

Since studies concerning speech applications in CAD are scarce, this means other office tasks (eg. text-editing) and industrial tasks (eg. parcel sorting) will also be included in the review. Such studies are usually comparative in nature, and are an important source of information for gauging the utility and problems of speech use in real-world applications. It should be noted that studies employing simulated speech input systems are not excluded; this will be indicated when a study is discussed.

This review will be in two parts. The first part reports on speech-manual input use in CAD, and the second part on non-CAD applications. The methodology and principal findings of each study will be presented.

2.8.1 CAD applications

Martin (1989) evaluated the utility of speech input in the context of a VLSI chip design package, and compared speech to typed, full-word input, single keypresses and mouse clicks. The study focused on two commonly-made claims about the utility of speech input that: (1) it is faster than typed input; and (2) it increases user productivity by providing an additional response channel. Seven graduate students took part in the study, and the data were derived from four subjects who had used at least one computer-based package for designing VLSI layouts.

Using a graphics design package called MAGIC (comprising 60 commands), subjects entered commands by (1) pressing a mouse button, (2) typing a full-word command on the keyboard, or (3) typing a single-key abbreviation or 'macro' for the command. The speech recogniser was a VOTAN system with a head-mounted microphone which could be activated or turned off with a voice command or a switch box near the keyboard. User histories were recorded on videotape; a split screen image of the graphics screen and monochrome text screen were captured. Each subject completed three phases: the first phase consisted of training, the second phase involved performing two structured tasks, and the third phase two longer real design tasks.

The results supported the benefits of speech input over typed, full-word commands, and to a lesser extent, over single keypresses. Users were able to complete more tasks (62% vs. 38%) when speech input was available. For the restricted set of commands that could be accomplished with mouse clicks, speech input and mouse clicks were equally efficient. The data support the claim that speech is beneficial because it adds another response channel. Also, subjects spent less time looking at the keyboard when speech was available. The results are interpreted in terms of a general 'ease vs. expressiveness' guideline for assigning modalities to tasks in a user interface.

The study by Kato and Tsuruta (1981) was aimed at evaluating the performance of a DP-100 connected speech recognition system that they developed for commercial use. In this study, the recogniser was tested in a LSI artwork system. The motivation for this research stems from the fact that with manually operated input devices: (1) the CAD operator must often shift his sight from the display to the keyboard; and (2) to assure a certain level of input speed and accuracy, an operator must pay attention to the devices. Therefore, a voice-operated CAD system might help to resolve these problems. Three solutions were investigated: (1) using spoken, single-word commands to substitute the function and numeral keys; (2) using spoken commands consisting of phrases, hence the vocabulary was smaller than the first solution; and (3) using both speech and manual input for multimodal interaction.

Four subjects edited a CAD drawing using 60 commands, and measurements of their performance were recorded in terms of (1) speed of task accomplishment, (2) error rate and (3) subjective preference. The results are derived from preliminary investigations of the first type of system which showed that speech input improved the speed and accuracy of task performance over keyboard input, and was preferred by users, particularly when they had to attend to the graphics display.

This preliminary study is weak in a number of ways. First, neither the CAD task nor the skills that are required to accomplish it is well defined. Second, and related to the first, the basis for measuring accuracy is unclear. Third, the study was biased, and so admitted by the investigators, in that the graphics tablet was used with speech input for data entry and interactive operation, but in the keyboard condition, the latter was the only input device for both types of interaction. Lastly, the lack of discussion on the implications of the findings on CAD performance makes it difficult to draw definite conclusions.

A final study involving the use of speech input in CAD is by Shutoh, Tsuruta, Kawai and Shutoh (1984). The CAD system, CGDS, included artwork generation in integrated circuit (IC) manufacturing. The speech recogniser was a speaker-dependent NEC DP-200. To evaluate the effectiveness of speech input, the following factors were examined: (1) recogniser acceptability; (2) operation capability; and (3) subjective preference.

The first experiment used 14 subjects (10 naive and 4 experienced users) who trained 50 commands once in Japanese. Subjects then repeated the words to check for recognition and the results indicated that experience was not an important factor in speech recognition. A second experiment compared speech input with keyboard and graphics tablet. Five subjects (3 experienced and 2 naive) were required to perform two types of tasks: (1) inputting coordinates to a circuit diagram; and (2) editing an IC mask pattern data. Measurements were taken of elapsed time and errors made in each task. Based on mean differences between naive and experienced, speech input performance was the same for both groups. But there appear to be some differences in mean performance within and between groups for keyboard and tablet input. Since no statistical tests on the data are reported, this result should be treated with great caution.

A study by Rickett (1987) which investigated the use of speech input for performing CAD tasks using FEMVIEW CAD deserves brief mention here. The study collected data from three sources: (1) the analysis of responses to a multi-user survey of end user attitudes; (2) behavioural performance measures from students learning to use the software; and (3) cognitive and affective data from experiments involving experienced users of the CAD package. A user model based on personality traits was developed using data from the experiments. Unfortunately, details of the methodology and evaluation techniques are not reported. The significance of the study was to demonstrate: (1) the practical problems of implementing speech recognition technologies in commercial software; (2) the development of a personalised user model which accounts for individual's idiosyncracies; and (3) the methods for applying simple evaluation techniques in order to assess software 'usability'. Rickett (1987), however, claimed that results from the experimentation to justify the model were inconclusive.

The next section examines selected studies that apply speech and manual input in text-editing, parcel and baggage sorting. The method aspects of these studies will be reported in lesser detail.

2.8.2 Non-CAD applications

In a text-editing environment, Gould and Alfaro (1984) conducted simulations of speech input, comparing editing performance of marked-up manuscripts using: (1) a simulated handwriting-recognition system; (2) a simulated speech-recognition system; (3) a full-screen text editor (XEDIT) and (4) a formatted version of XEDIT. Twelve volunteers (4 secretaries and 8 principals) took part in the experiment. All had used the text editor regularly. The results showed that editing was done much faster with handwriting and speech input systems (50% and 90%, respectively) than with a text editor despite their experience with the editor.

The findings must be qualified in terms of the simulations used. Unlike real recognition systems, (1) there were no user interface problems (eg. difficult or cumbersome to use commands); (2) there was no need for any special editing conventions; (3) there was no waiting for system response time; and (4) perfect recognition was assumed. Also, unlike the text editor, users could concentrate on the manuscript, hence attention was not divided between the manuscript and another display. "This spatial displacement and division of attention are time-consuming and distracting, in part because they lead to visual search and forgetting" (Gould & Alfaro, 1984, p. 404). The above factors might have contributed to making writing and speaking faster than Xedit-editing.

Morrison, Green, Shaw and Payne (1984) examined the effects of input modality and of command structure on text editing. The speech-driven editor employed a Heuristics H2000 SpeechLink speech recogniser (see Green et al., 1983) which uses an adaptive algorithm to compensate for changes in diction during the task. Performance measures and satisfaction ratings were obtained from 20 subjects (10 skilled typists and 10 non-typists) using two different designs of editor, one requiring more but simpler commands ('short transactions'), the other needing fewer but more complex commands ('long transactions'). Each subject used the same editor in two versions, one with all input from the keyboard (ie. keyboard-only mode), the other with spoken commands but typed parameter strings (ie. speech-plus-keyboard mode).

The results indicate that short transactions were preferred, although they were not always error-free. Speech input was consistently rated lower than keyboard input by typists;

non-typists initially preferred speech input but changed to preferring keyboard input when they gained more practice. The dislike of speech input may have been due to: (1) the limited hardware; and (2) the switching between modalities during a command which was inherently disruptive to both skilled and unskilled keyboard users.

In a multi-comparison of input devices, Haller, Mutschler and Voss (1984) investigated the performance of a graphics tablet, speech recogniser, cursor keys, lightpen, mouse and trackball for positioning the cursor and correcting typing errors. The speech recogniser was a speaker-dependent CSE 1060 (Computer Gesellschaft Konstanz) system fitted with a headset microphone. Six subjects with computer experience participated in the experiment. Two types of measures were obtained, positioning and replacing times. The results showed that speech input of location, that is, by speaking the coordinate values, was slower than all other input devices above. Comparing speech input with keyboard for correcting typing errors, keyboard entry was slightly faster, and less error prone than speech. A ranking of the devices showed the lightpen to be a quicker and easier to use device than the graphics tablet and speech recogniser.

Leggett and Williams (1984) conducted an experiment to evaluate speech and keyboard input as a data entry mode for computer programming. Twenty-four university students entered and edited segments of program code; measures of speed, accuracy and efficiency were used to compare the two input modes. The subjects were able to finish more tasks using a keyboard, as opposed to speech (70% compared to 50-55%, respectively). However, keyboard entry produced a higher error rate than speech, despite the subjects being more experienced with keyboard input than speech. Speech recognition error rates were 17% for the input task and 16% for the edit task. The act of stressing a word to help the system was a notable problem with speech use. Other problems included speaking too softly, lip-smacking and failing to maintain the position of the microphone. All of these factors contributed to the low speech recognition rate.

The use of an ASR system in television subtitling was investigated by Damper, Lambourne and Guy (1984), using the NEWFOR system - a computerised subtitling aid which operates in conjunction with the ORACLE service. The speech recogniser was an Interstate VRT-300, a 100-word, speaker-dependent isolated word device. The vocabulary consisted of 25 words for selecting colour, position and size of each subtitle frame. Measures of input rate and errors were obtained from 7 subjects. The results showed that speech input increased preparation time by 9%. But it significantly reduced the time spent transferring between text and style entry. The error rate for style entry by keypad was consistent and averaged 5%, whereas speech errors ranged from 5-15%. The speech errors were partly due to: (1) dual-task interference (ie. speaking commands concurrently with entering text and style); (2) drifts in the task voice; and

(3) noise from depressing keys on the keyboard. This study indicates the difficulty of using speech and manual input simultaneously when the manual task is complex.

Johnson, Long and Visick (1985) experimentally compared four input devices in a data entry task in order to identify a suitable device for parcel sorting. Three different keyboard layouts (QWERTY, Matrix and Parcel Sorting Mechanism or PSM) and a speech recogniser were tested. This review, however, will focus on comparisons between the recogniser and the QWERTY keyboard since the latter will be used in the thesis. The speech recogniser was an Interstate VOTERM II, fitted with a SHURE SM10 noise-cancelling microphone and headset. It is a speaker-dependent, isolated-word system with a vocabulary of 100 words. The keyboard was laid out in the standard alphanumeric configuration but without the space bar, punctuation and function keys. An analysis of learning was made on the basis of percentage errors and task completion time from 12 subjects. The QWERTY group of subjects were all female touch typists.

The results showed that when users' hands were busy at the sorting task, speech yielded a 37% improvement in entry time; but it also produced an error rate of 40-80%, compared to the keyboard error rate of 10%. This improvement in efficiency disappeared when the 'hands-busy' component of the task was removed. That is, the advantage for speech disappeared when users simply read or keyed-in the destination names from a list, as opposed to reading the names from the parcels and sorting them. This study suggests that speech has value in providing an additional response channel in situations where the users' hands are likely to be busy at other tasks. The QWERTY keyboard required typing and coding skills which makes it a less favourable device since most parcel sorters do not possess such skills.

In a similar task environment, Frankish, Jones, Madden, Waight and Stoddart (1987) compared two types of vocabulary in a simulated parcel sorting task using speech input. One type used place names drawn from operational use, another type used alphanumeric codes based on the International Civil Aviation Organisation (ICAO) alphabet. The speech recogniser was a Kurzweil Voicesystem, a speaker-dependent, isolated-word with a vocabulary of 1000 words, and a claimed accuracy of 95% or better. Sixteen subjects, all members of the general public, were evaluated on their task performance in terms of both speed (task completion time) and accuracy.

The results showed the recognition rate for ICAO alphabet and place names was 78% and 82%, respectively. The ICAO alphabet was claimed to be superior in that the proportion of substitution errors which were detectable was substantially higher (98%) than for place names (88%). The findings suggest the importance of vocabulary design in developing a practical system for parcel sorting.

Two other studies which will be reported in brief are by Poock (1982) and Nye (1982). Poock (1982) compared the speed and accuracy of speech and typed entry of commands and control inputs (eg. logging into different host computers, reading messages, etc.). Speech input was 17% faster than typing; and typing produced 183% more errors than speech. Also, users preferred speech over typing. Nye (1982) compared speech to keyed entry of locations in an airline baggage sorting task. The keyed entries were three-digit codes for different flight numbers, the spoken entries were the destination cities. With the keyed input method, errors ranged from 10% to 40%; with the speech input method, errors were reduced to less than 1%. But in a separate experiment which used similar data type (ie. three-digit codes) for speech and keyboard input, Nye (1982) found speech input capabilities to be less successful. It became useful when users were required to enter the names of cities. This suggests that the nature of what information users enter into a system is crucial in speech use.

2.8.3 Conclusion

As these studies show, the results of comparisons between speech and manual input (particularly the keyboard) are often contradictory and ambiguous. In some cases, speech input yields faster, more efficient entry; in other cases it does not. The quality (and cost) of the speech recognition technology used may be one factor responsible for the variable results. With very high quality speech recognition systems, used in an optimal environment, speech input often yields low error rates relative to keyboard input; whereas the reverse occurs in other situations. Hence, these results do not verify the claim that speech is a faster human response channel than typing.

However, the results suggest that: (1) speech input capabilities may not simply make the existing form of interactions more efficient; rather speech may change the nature of *what* information users enter into a system; and (2) speech input has potential in complementing manual input devices. It is important, however, to consider carefully which particular manual input devices can combine well with speech input and for which particular aspects of task. These questions will be addressed in this thesis.

2.9 SUMMARY

This chapter describes the technologies of speech recognition and manual input and their applications. In certain applications (eg. manufacturing and parcel sorting), speech input already has advantages over manual input. The use of speech input, however, is still largely experimental as is evident from the studies reported here. This is partly because of the ergonomic constraints related to its use. Given the constraints, successful applications of speech input may be achieved if it is supplemented by, and integrated with, other manual input devices. The choice of a manual input device will depend on the task and the characteristics of the device. The graphics tablet has the properties of an indirect and direct device, hence it is a multi-functional input device, suitable for supporting all three functions mentioned above - information input, selection and positioning plus moving of data.

NEXT CHAPTER HIGHLIGHTS

Chapter 3 describes CAD technology in terms of hardware and software developments and the impact of CAD in the workplace. The important role of input devices in CAD performance is further emphasised through description of CAD tasks. The relationship between CAD and the design process is discussed to contextualise CAD within the overall design environment.

CHAPTER 3

Background to CAD Technology and the Design Process

Overview

3.1 Introduction

3.2 What is design and the design process?

3.2.1 Empirical studies of design process

3.2.2 Conclusion

3.3 What is CAD?

3.3.1 Why use CAD?

3.3.2 Constraints of CAD use

3.3.3 Conclusion

3.4 CAD/CAM development

3.4.1 A brief history of CAD/CAM

3.4.2 CAD/CAM development in advanced and developing countries

3.5 CAD systems

3.5.1 Hardware

3.5.2 Software

3.5.3 CAD tasks

3.5.4 Conclusion

3.6 Human factors in CAD

3.6.1 Commands

3.6.2 Graphical and alphanumeric data

3.6.3 System feedback

3.6.4 Conclusion

3.7 Empirical studies on human factors in CAD

3.8 Summary

Next chapter highlights

CHAPTER 3

Background to CAD Technology and the Design Process

OVERVIEW

This chapter introduces hardware and software components of CAD systems. The aim is: (1) to identify some human factors issues that need to be considered in configuring a demonstrator CAD system; and (2) to familiarise the reader with the relevant aspects of CAD technology. CAD is a part of the design process. Therefore, some knowledge of design is required. A brief introduction to the design process is presented first, to contextualise CAD within the design environment.

3.1 INTRODUCTION

The introduction of CAD systems is a part of a broader pattern of socio-technical change (Eason, 1988; Cross, 1984). The rapid pace of change in computing, allied to rapid changes in other branches of technology (eg. manufacturing), results in changes in the way people work, learn or even spend their leisure time. To accommodate such socio-technical changes, design processes and practices change as well. The following sections will trace some of the changes in the design process, including what is meant here by design.

3.2 WHAT IS DESIGN AND THE DESIGN PROCESS?

The term *design* does not have a single meaning. As Nickerson (1986) explains, sometimes it connotes a process, or the product of a process, that precedes an attempt to build something. The purpose of a design, in this sense, is to specify the characteristics of the thing to be built. At other times the term connotes simply the characteristics of something that has been developed. Whether the design in this case preceded the thing to which the design refers is incidental.

Whitefield (1986b) defines design as *the creation of specifications to construct objects that satisfy particular requirements*. In any design process the three major functions are: synthesis, analysis and presentation (Majchrzak et al., 1987). CAD is a technology for supporting the design process. This means, as a design tool, CAD assists the designer in the creation, modification, presentation and analysis of a design.

To contextualise CAD within the design environment, it is necessary to provide an overview of design and the design process. Carroll and Rosson (1985) have suggested four aspects of design that essentially characterise it:

- design is a *process*, it is not a state and cannot be adequately represented statically;
- the design process is *non-hierarchical*, neither strictly bottom-up nor top-down;
- the process is radically *transformational*, involving the development of partial and interim solutions which may ultimately play no role in the design; and
- design intrinsically involves the *discovery of new goals*.

Carroll and Rosson are trying to convey that the design of a system evolves throughout the design and development process, thus a system is not simply specified and built. At the beginning of the design process some of the low-level goals are known as well as some high-level goals. Throughout the design process, through compromises and tradeoffs, these goals build into a more complete and coherent picture, as goals are added, changed or discarded. Solutions to design problems often require creativity, and consequently, the process of design cannot be described as completely rational or logical.

In addition to the above characterisation, John (1988) describes design as: (1) often reflecting the needs, values, and purposes of designers in orders and patterns that give meaning; (2) a multi-disciplinary activity demanding communication boundaries; and (3) an activity which sits in the context of organisational, social, political, economic and scientific constraints, incorporating values from, imparting them to and reinforcing them within such. This characterisation emphasises the importance of design management and coordination (involving managerial, marketing, technical, etc. functions) within an organisation. The goal is to create design awareness and to achieve greater competitiveness (Ughanwa, 1988; Eason, 1988).

Nadler (1989) identified three common assumptions of conventional design processes which can be fallacious. First, that all problems are alike and can be approached in the same way. Yet the engineering profession alone commonly recognises at least four generic sorts of problems. These are: (1) improvement of an existing system; (2) diagnosis and remedy; (3) development of a new system; and (4) development of a new use for an existing system. These different sorts of problems clearly require different approaches and methodologies.

A second assumption is that just knowing about the latest technologies and analysis techniques will produce solutions. But the first solution that emerges may not be the best one, and when implemented may turn out to be lacking in desired characteristics of low cost, high quality and low process time. A third assumption is that because the research approach (eg. fact gathering, model making, etc.) is appropriate in some situations it must be equally applicable to others. But real-life problems in planning and design often turn out to be ill-structured problems than the patterned research approach (Nadler, 1989).

A brief review of the historical development of the design process suggests that there are several approaches to design (John, 1988; Rouse & Boff, 1988; Cross, 1984; Jones, 1970). The approach of the 1960s was prescriptive and systematic (or *structured methods*), borrowing what was seen as scientific methodology from systems engineering techniques of military and space programmes. The main design stages have been generally described as analysis, synthesis and evaluation. Proponents of this approach support the maxim that 'designer knows best' (eg. Hubka, 1983).

A second approach prescribed *participative* design, that is, users participate with designers in the design process. This approach arose out of disillusionment with the structured methods, maintaining that one cannot understand the problem without a concurrent concept of a possible solution. Thus, design was not so sequentially ordered. Proponents of this approach claim that the 'user knows best' (eg. Rittel, 1984). A third approach was based on Popper's conjecture/refutation model, in which the designer raises conjectures until an irrefutable solution is attained. This process is highly iterative and interactive, hence called *interactive design*. Proponents of this approach include Broadbent (1979).

In order to understand better the design process, the next section looks at some empirical studies.

3.2.1 Empirical studies of design process

This section is not intended to be a detailed review of design studies, but an overview of the main findings derived from some investigations. The aim is to highlight some important characterisations of the design process. Selection of these examples is based on the criteria that the studies are empirically-based, and are involved in design analysis with and/or without CAD use. This second criterion is crucial to an understanding of the role of CAD in the design process. For a full review of design studies, see Rouse and Boff (1988), Whitefield (1986b) and Lera (1983).

Several investigators have observed real designers working on real design problems (eg. Whitefield & Warren, 1989; Tovey, 1989; Eckersley, 1988; Ballay, 1988; Hammond, Jorgensen, MacLean, Barnard & Long, 1983; Carroll, Thomas & Malhotra, 1980). Their research into design has tended to produce results that generally agree with the characteristics outlined by Carroll and Rosson (1985).

Hammond et al. (1983), for example, interviewed experienced system designers in an attempt to determine what types of decisions they made during the design process and on what these decisions were based. Their findings are difficult to summarise, but the quoted comments of the designers suggest that designers' 'theories' of users tended to consist of broad generalisations about user behaviour. There did not appear to be any recognition that user behaviour might vary across tasks. Designers, however, did recognise that their knowledge of users are inadequate, and that they were largely unaware of resources they could tap to resolve them.

Thomas and Carroll (1984) summarised some of the findings of a number of design studies conducted at IBM's T.J. Watson Research Center as follows: (1) design problems seemed structured in terms of subproblems; (2) the subproblems were dynamically produced during design, not completely specifiable at the beginning; (3) designing in space seemed easier than designing equivalent problems in time; (4) a crucial aspect of design is specifying goals; and (5) goals stated in high-level terms were not interpreted identically, even by experts in a field.

In trying to understand how the design process is affected by the use of CAD systems, Ballay (1988) observed designers solving an industrial design problem using traditional methods and CAD systems. Comparing a model of the design process with the products of designers' problem solving, he claimed that the early stages of the process were particularly important to the production of successful designs. The early stages were defined from three viewpoints - design as an ill-defined construction task, design as a 'visual task', and design as a series of information transactions. The visual aspect of design means that CAD contributes to the early stages of the design process via several forms of representations - perspective drawings, notations, dimensions, matrices, orthographic projections, solid models and procedural representations.

But Rooney and Steadman (1987) claimed the main contribution of CAD as being to the later stages of the design process. They see such CAD techniques as solid and surface modelling, finite element analysis, etc. (see Section 3.5.2) as being part of the evaluation stage of the process. Further, they locate other CAD techniques such as automated draughting and the electronic transmission of the final design to numerically-controlled (NC) machines and robots as the last phase of the process, manufacture.

In an analysis of design drawing and CAD in industrial design process, Tovey (1989) found CAD: (1) to be inherently unsuitable for innovative design; (2) having potential for contributing to evolutionary design; (3) inhibiting fluid design thinking and design modelling; and (4) supporting design styling, evaluation and integration. Whitefield and Warren (1989) in their investigations of designers' behaviour during the design process, found CAD designers recruiting more operating knowledge, largely at the expense of evaluative domain knowledge, than drawing board designers. This operating knowledge involves knowing how to 'navigate' the drawing, and the management and production of the drawing.

The above findings raise a number of questions about how CAD systems relate to the design process, and how future systems should be designed. This was a central concern of Pikaar's (1989) study, aimed at identifying how design tasks should be allocated to CAD systems. A situation analysis, involving a formal system description and a reconstruction interview, was carried out amongst drawing board designers. The findings showed that the designers: (1) liked the initial design phase; (2) disliked the drawing activities and the documentation phase; and (3) preferred to work with two or more drawings on the drawing board. A global allocation of system tasks to the design of CAD systems was thus recommended. Further discussion on the possible impact of CAD on designer behaviour and performance will be made in Section 3.7.

3.2.2 Conclusion

Whatever the methodology employed in these studies (eg. verbal protocol analysis, situation analysis), one common finding that supports the many views on design is that: design is a group of related but distinct activities. Essential to these activities is the design problem information - the information that gives substance to the problem and guides the designer in his or her efforts to solve it. The character of this information constrains the nature of the solutions that are possible. CAD has proved to be unsuitable for innovative design and tends to inhibit design thinking, but it is found to be more effective in supporting design styling, evaluation and integration. The next section takes a closer look at CAD, its technology and its impact on industrial development.

3.3 WHAT IS CAD?

As mentioned earlier, a design process involves synthesis, analysis and presentation. A CAD process is iterative and embedded in a design environment which requires that synthesis operations be associated with the human designer and analysis operations with the computer. Within these capabilities lie the two contrasting definitions of CAD: on the one hand, the electronic drawing board, and on the other, the 'intelligent' conceptual design tool which accommodates design, production and manufacturing.

In a non-CAD system (or traditional design process), the design functions are carried out manually with the aid of draughting instruments, calculators, handbooks, etc. But a CAD system substitutes a computer and I/O devices for these traditional aids. Hence, a CAD system is defined here as *an integrated configuration of design software and hardware which utilises* *interactive computer graphics to support design activity.*

The use of computers for word processing is commonplace. There are many similarities between word processing and CAD, in that a word processor allows an operator to create, manipulate and store text, CAD allows the creation, manipulation and storage of drawings. The standard input device in both applications is the alphanumeric keyboard for performing the functions described in Chapter 2, and the output device is the alphanumeric screen for displaying system information.

However, there are basic differences between the two. First, CAD requires a graphics screen of reasonable resolution and size, and colour if the drawings are complex for interpretation on a monochrome screen. Second, CAD requires a pointing device for graphical input. The choice of an appropriate device will be determined by its accuracy, resolution and portability (see Chapter 2). The graphics tablet often complements the keyboard in CAD use. Third, CAD requires a plotter for putting the drawings on paper. The choice of a plotter will depend on its accuracy, resolution, speed, number of pens and maximum drawing size. In some instances, CAD requires a drawing scanner which scans existing paper drawings directly into the CAD system.

The differences are more obvious in terms of the software requirements for operating CAD systems. Generally, CAD software is more complex and demanding in terms of its underlying data structures and processing activities than word processing (Lang, 1985). In the domain of mechanical engineering alone, there are over 150 general-purpose software products, falling into the category of two dimensional (2D) design and draughting (ie. CADD), with little emphasis on the first 'D' and a lot on the second (Owen, 1988). Ironically, it is often said that CAD has evolved from design into computer aided draughting (Billsdon, 1987).

Basically, CAD systems provide facilities to add, modify, or delete: (1) straight lines usually in a variety of line-styles (eg. solid, dashed, dotted, etc.); (2) circles and arcs, also in a variety of line-styles; (3) text usually in a variety of fonts or styles at any size or angle; (4) symbols, a collection of lines and text to form shapes which are required many times in one or more drawings (eg. valves, pumps, etc); and (5) attribute data, ie. textual data which is logically associated with a line, text or symbol in a drawing.

In short, CAD systems could be used to produce any drawing, but generally they are best at producing drawings which: (1) contain a lot of repetitive detail (eg. architectural drawings); (2) need to be of a very high quality (eg. quotation drawings); (3) require repeated modification (eg. drawings of prototypes or products under development); (4) require analysis and calculations (eg. mechanical drawings); and (5) the data can be reused for several applications (eg. circuit schematic drawings) (Billsdon, 1987). Other potential benefits of CAD use and their constraints are described below.

3.3.1 Why use CAD?

With CAD, design updates can be done easily and the final hardcopy output presented neatly. Thus, design productivity is greatly increased (Majchrzak et al., 1987; Rooney & Steadman, 1987). Before the computer era, drawings were done with pencil and paper. It is a painstaking process to prepare a technical/engineering drawing and modify it. Long lead times for drawing preparation were accepted when designing manually. Inevitably there were the potential human errors in interpreting not only the sketches but also the final drawings. This situation changed when CAD systems were introduced to substitute for the drawing board and instruments, as generally claimed about CAD.

Other benefits of using CAD systems often cited in the literature (eg. Pikaar, 1989; Tovey, 1989; Billsdon, 1987; Senker & Arnold, 1984) include: (1) reduced product development lead times; (2) faster turn-round quotations; (3) greater design rationalization; (4) enhanced market image; (5) better design; and (6) improved communications with customers. Due to these likely benefits many companies are willing to invest in CAD/CAM systems (see Goodwin, 1986). But the introduction of CAD systems into industrial organisations has not always been successful. The next section discusses some of the constraints involved.

3.3.2 Constraints of CAD use

Evidence exists (eg. Haase, 1989; Rose, 1988; Roth, 1988; Majchrzak, Collins & Mandeville, 1986; Rzevski, 1984; Begg, 1984; Leesley, 1978) that the reasons for observed adverse effects of CAD use may be traced to the incompetent management of the change and superficial problem analysis. For example, a recent review of CAD (PC Week, August 1, 1989, p.13) illustrates that:

> "It may come as a surprise to draughtspeople who are still working in an unautomated office that it takes roughly the same time to create a drawing on paper as it does to create it on a computer, the difference is error changing and productivity."

Incorrect design of CAD systems has on some occasions caused de-skilling of design work. Although "in fact, what CAD systems do is to remove drudgery and frustration and to improve product quality by allowing the designer time to be more precise and thorough" (Begg, 1984, p. 16). Incorrect selection of CAD systems has sometimes resulted in the imposition of unreasonable constraints upon the freedom of design decision making. These problems are well summed by Roth (1988) in an article entitled *Is the Master Serving the Slave?* He commented:

"Those who claim that CAD equals better design do the technology disservice by raising false expectations... computers cannot 'design' because they cannot think - they impose order. CAD systems are complicated, unfamiliar, and distort the design process of idea, concept, analysis and finally, detail. Thus, the designer is forced to adapt to the system, and must become an enthusiast to use it effectively".

Changing from 2D CAD to three-dimensional (3D) systems, which allow one to look at an object at different angles, has its problems. The vast majority of 3D CAD installations increase productivity. Just as important, they also provide design and engineering capabilities previously unavailable. But the time required to reach '1-to-1 parity' (or ramp-up time), that is, the hours required before a new user's output equals or exceeds what s/he could do without the package, is considerable (Haase, 1989). "There are three components to becoming productive in 3D CAD - the hardware and software costs, the cost of training the person, and the cost of tweaking the software to match the company's use" (Slinn, PC Week, July 18, 1989, p.16). He recommended that a company should allocate between 100 and 200 percent of the cost of the software for training. He added that it may take 100 or more hours after the training before reaching parity. A spokesman for Autodesk, the maker of AutoCAD, confirmed that training and lost time before a user becomes truly productive is a big issue. "The micro-based 3D CAD package has become a sophisticated package. But vendors don't have the market to provide training and support like they do on a mainframe system" (PC Week, July 18, 1989, p.17).

The fact that CAD systems are becoming more complex, has made the transfer of design information both within the organisation and to other manufacturers a problem for many organisations (Rose, 1988). "Now that we have got computers we are being forced to think about what it is that we want to transmit, and the best way of doing it... obviously we can just produce more drawings... but that is only an intermediate solution" (Pearson, 1988, p. 66). The problems are complicated by the speed at which CAD systems are evolving. As Pipes (1986, p. 51) explained,

"...too much software is churned out with little relation to demand. Vendors prefer to develop in width, not depth, so as to satisfy as many applications or potential markets as they possibly can. Systems tend to be fragmented, flawed, idiosyncratic and inevitably compromised in their design".

A survey conducted for the Engineering Council by Cranfield Institute of Technology, UK, concluded that over a third of users were disappointed in the return on their CAD/CAM (computer aided design/computer aided manufacture) investment. This has been reinforced by a survey of 200 users by consultant Bryar & Gaskell that 73% of the UK's CAD/CAM systems are only reaching 53% of their potential throughput (source: Computer Weekly, November 13, 1986, p. 51).

In an effort to understand how CAD can be optimised in an organisational setting, Majchrzak et al. (1986) compared the attitudes of CAD and non-CAD users about their jobs and workplace. The results suggest that: (1) CAD introduces standardisation and this tends to be viewed negatively by users; (2) CAD implementation does not guarantee an increase in creative, nonroutine activities for all CAD jobs; (3) CAD does yield an increased responsibility among users for the entire design process; and (4) the increased coordination activities found with CAD do not necessarily translate to an increased reciprocal interdependence.

3.3.3 Conclusion

The above excerpts and findings suggest that the slow up-take of CAD into certain design areas is due to a number of barriers, many of which are human factors issues, in particular training, ease of use and system design. Allied to these are economic and organisational barriers, in particularly system costs and the management of change. Companies who try to justify CAD on the basis of unrealistic productivity estimates not only face major problems in proving that the investment is viable but they also create major industrial problems. Thus, there is little evidence that even after 25 years or so of application, CAD/CAM is used much at the initial stages of design. The problem is that the appreciation and application of CAD/CAM has always lagged behind the development of the technology. The next section will trace the historical development of CAD/CAM technology.

3.4 CAD/CAM development

This section is in two parts. The first part looks briefly at CAD development, and related technologies of CAM (computer aided manufacturing) and CIM (computer integrated manufacturing). The second part examines CAD/CAM development in advanced and developing countries - its impact, problems and future. The purpose is to afford an understanding of how CAD/CAM originates and how it is used today in different parts of the world.

3.4.1 A brief history of CAD/CAM

The development of CAD dates back to the early 1960s when cathode ray tubes (CRT) and light pens were used in the USA's SAGE system. The introduction of the SKETCHPAD by Sutherland illustrated the potential use of the CRT system as an electronic drawing board. Shortly after interactive computer graphics became available in the 1960s, 2D draughting, 3D and wireframe CAD systems were developed, led by the aerospace and automobile industries (eg. McDonnell Douglas, Lockheed and Ford).

In the 1970s, with the drastic drop in computer software and hardware costs, increases in their capability, and the development of easier to use software, CAD became more widely accepted. During this period, software for 3D solid modelling and shaded colour graphics were developed. Many software programs for engineering analysis applications were also integrated into CAD systems. At this time CAD was moving from computer-aided draughting to computer-aided design and modelling (Majchrzak et al., 1987).

The continuing development of super-mini and microcomputers added a new dimension to CAD in the 1980s. The development of CAD software for personal computers led to the widespread use of CAD in the offices and homes. For example, the worldwide sale of AutoCAD software has now exceeded 230,000 copies, with over 75,000 copies installed in Europe (source: First Draft, July 1989, p. 10). Developments in artificial intelligence techniques spawned interest in the development of expert CAD systems (eg. Rosenman, Gero, Hutchinson & Oxman, 1987). Ideally, an expert CAD system can automate the entire design process (Begg, 1984; Majchrzak et al., 1987), but this is not yet a reality.

The exploitation of CAD by manufacturing industry illustrates the widespread adoption of computer technology within an application domain. The electronics manufacturing industry was quick to adopt the opportunity afforded by CAD to reduce the cost of printed circuit board (PCB) and integrated circuit (IC) design (Heap, 1986). The construction industry was next, and has continued to be an increasing user of CAD. The European CAD scene in 1987 showed CAD application as comprising of: 59% mechanical computer aided engineering (MCAE); 22% electrical and electronic design (IC/PCB); 13% architectural, engineering and construction (AEC); and 9% mapping (source: CADCAM International, June 1987, p. 45).

The use of CAM machinery, such as computerised numerically controlled (CNC) machine tools (based on NC technology developed in the 1950s) and software controlled robots in manufacturing operations, has led to advances not only in the CAD/CAM technology but in other related technologies as well, such as CIM. CAD is only one part of these technologies. In essence, CAM refers generally to the automation of manufacturing processes (eg. tool grinding) and related activities in the manufacturing environment (eg. inventory control, quality control, etc.). CIM refers to the linking of the various islands of automation in a manufacturing plant, for example, the administrative/ planning section and the factory floor section. CIM is now seen as the goal of modern manufacturing.

The design of CIM systems, especially as fully automated systems (the so-called 'peopleless' option) has led to renewed interests in the fundamental principles of allocating functions between machines and humans (see Clegg, Ravden, Corbett & Johnson, 1989). The dominant trend in systems design follows a sequential technology-driven approach, whereby the designers automate as many functions as possible, and the humans are allocated whatever functions remain. Any behavioural and/or organisational requirements are rarely, if ever, made explicit at this early stage of the design process (Meister, 1987).

As mentioned earlier (Section 3.2), technological development should be a part of a company's overall strategy, and achieves its maximum potential when used as a catalyst for wider organisational change (Clegg et al., 1989; Edwards, 1989; Majchrzak, et al., 1986; Schaffitzel & Kersten, 1985). For example, it is believed that the success of Japanese manufacturing systems stems as much from their organisation innovations as from their new technologies (Clegg, 1986).

3.4.2 CAD/CAM development in advanced and developing countries

Several reviews of CAD/CAM development exist in the CAD literature (eg. Majchrzak et al. 1987; Whitefield, 1986b; Groover & Zimmers, 1984; Begg, 1984; Simon, 1982). These reviews have tended to focus primarily on CAD development in advanced industrial nations with almost no mention of developments in the developing industrialised countries of the 'third' world. While most developing countries (eg. African countries) might never have the chance in the forseeable future to catch up technologically, Asian countries such as South Korea, Taiwan, Hong Kong, Singapore and Malaysia (known as *industrialised threshold* countries) are increasingly using CAD in the production of labour-intensive products, such as semiconductors, circuits, accumulators and other electronic components.

This short introduction to CAD/CAM development will examine developments in both parts of the world, taking United Kingdom (UK) and Malaysia as examples of an advanced and a developing industrialised country, respectively. The aim is to assess the impact of CAD and/or CAM technology in terms of acceptance, applications, problems and future developments. Space precludes giving much detail, thus the description below is an overview of the general situation.

Advanced nation: the case of the United Kingdom

In the UK, the impact of CAD/CAM initially lagged behind technological advancement, partly due to draughtsmen being unwilling to adopt new technology that makes inappropriate demands on the user (Simon, 1982). The fear of being made redundant was a real one. But the hesitant approach to CAD/CAM possibilities changed, in part, as a consequence of two diverse issues. Firstly, technological advances in the fields of computing and electronics, resulting in cheaper systems and a wide variety of applications. Secondly, a decline in the manufacturing output between 1978 and 1986. Concern about UK's poor industrial performance in world manufacturing exports when compared with their major foreign competitors (eg. US, Japan, Germany) has led to a search for ways of promoting the competitiveness of the manufacturing industry, and in particular for ways of ensuring that the sector uses the latest and most advanced manufacturing technology (Finkelstein, 1988; Ughanwa, 1988).

In 1982, the Government developed a set of programmes which was designed to promote awareness of CAD/CAM. The latter was given an enlarged profile in both the popular and technical literature. In 1986, vendors of CIM equipment began financing a number of centres for CIM development and education (eg. IBM's Warwick Cell). The Department of Trade and Industry also established programmes to promote awareness of CIM. However, these are still in their infancy. The highest priority in UK is to train sufficient engineers and designers to use CAD/CAM (Computing, 1986). Universities and colleges have taken initiatives, along with many CIM centres around the country.

A national CAD/CAM survey which covered 129 companies in eight industry sectors (eg. computing/electronics; electrical, mechanical engineering; architecture; vehicle manufacturing; education; etc.) revealed that: (1) 45% of the respondents planned to increase their spend on CAD/CAM systems within a year; (2) workstations were their first priority, followed by applications software; and (3) the largest users of CAD/CAM were in computing and electronics (73%), next came civil engineering and architecture (55%). However, there was also growing concern amongst CAD users on the problem of inter-system communications and response time on their CAD systems (source: Computing, September 25, 1986, pp. 30-32).

In terms of application domains, the forecast for 1990, by Dataquest is: 45% MCAE; 35% PCB/IC; 14% AEC and 6% mapping (source: CADCAM International, June 1987, p. 47). On the European scene, it is predicted that the European CAD/CAM market will rise to £1,800 million in 1990s, accounting for about 30% of the world CAD market versus 50% for the US. Germany will supersede the UK, while France's user base will expand fast, as will Italy's. For details on the UK scene, see Finkelstein (1988), Ughanwa (1988), Whitefield (1986b) and Simon (1982).

Developing nation: the case of Malaysia

In Malaysia, and elsewhere in Asia (eg. Singapore, Hong Kong, India, Philippines, Indonesia, Thailand), the growth of CAD/CAM usage is proceeding at a much slower pace than western industrial nations, and the areas of implementation vary greatly. The factors contributing to the lack of uptake of CAD/CAM include: (1) the readiness of these countries to employ new technology due to economic and political reasons; (2) the pressure of competition; (3) computer illiteracy; and (4) the problems associated with the transfer of technology. Since systems are not adapted to the needs of the developing countries (eg. locally-oriented dialogue design), their availability did not help to speed up computer nor CAD appreciation. Thus, CAD/CAM systems have been employed mostly by multinationals, largely on engineering and construction projects.

Malaysia's response to CAD began in the 1980s, with initiatives by multinational companies (eg. Shin-Etsu Handotai (SEH), Rothmans) and local architectural firms (eg. Akitek Bina Jaya). Malaysia's huge proliferation of small manufacturing concerns simply could not entertain adopting CAD/CAM technology: (1) until its costs come down; and (2) until there is sufficient appreciation of the technology through education and training.

Unlike the UK, computer and CAD technologies are given much profile in the national newspapers (eg. The New Straits Times, The Malay Mail, The Star) in order to increase literacy on all aspects of the technology. The involvement of CAD vendors such as IBM (Malaysia) Ltd. and Hewlett-Packard, in providing educational programmes for executives help to push the technology further. Although committed to contributing to the nation's progress, IBM, in particular was faced with the problem of 'piracy'. Market sources estimated about 70% of all PC sales were dominated by pirate ones (source: Business Times, August 30, 1986, p.15). Inadequate exercise of the copyright laws discouraged vendors from making possible the transfer of technology. Ironically, because of the low and affordable prices of these pirated systems, more people were able to afford them, which led to an enhancement in computer awareness in general, and CAD in particular.

Malaysia continues to make a concerted effort to harness the benefits of IT and CAD/CAM technologies in an attempt to keep pace with other countries in the region. This effort is witnessed by the fact that science and technology has been integrated into national development planning for the current Fifth Malaysia Plan and the setting up of a National Science and Technology Information Centre to aid research and development (Business Times, August 30, 1986). Unlike the UK, the highest priority is to train the educationists and implementors of the technology, rather than potential designers. For example, the Education Ministry has conducted in-service training for teachers involved in the Computer Literacy Pilot Project. Universities and polytechnics have initiated 'hands-on' CAD training for all engineering students as a serious move towards implementing the technology.

Malaysia's continued reliance on the electronics industry for its revenue is expected to increase the use of CAD/CAM, as spelt out by the Industrial Master Plan. The industry is the main thrust of the country's manufacturing sector, with an output of some 16.7 billion ringgit (£3.7 billion) by 1995 (source: Malaysian Business, May 1, 1986). Despite being the third largest producer and leading exporter of semiconducters in the world, Malaysia's involvement in the Japanese, American and European dominated microprocessor industry, was mainly in the production and assembly of the product rather than in the design itself. This situation changed with intervention from the Government as an initiative towards implementing a one-to-one transfer of technology.

The introduction of CIM in multinational manufacturing companies (eg. SEH), and in the national car industry (ie. PROTON) has helped to boost the implementation of CIM in Malaysia. A number of companies have carried out various feasibility studies, with help from international management consultants (eg. Arthur Anderson). CAD, in particular, has potential in the fast expanding car industry for designing models of Proton Saga cars - a Malaysian product using Japanese technology. For details on the Malaysian scene, see Malaysian Business (1986). The Asian Computer Monthly (1986) provide details on the different responses to CAD/CAM in other Asian countries.

In conclusion, the uneven acceptance of CAD/CAM is not surprising, particularly in developing countries. In the UK, the take-up of CAD/CAM to increase design and production efficiency has been much more dramatic than in Malaysia. In the case of Malaysia, there were more barriers to acceptance, ranging from computer and CAD illiteracy to technological adaptation. The above comparisons also illustrate the significant roles played by computer vendors and government bodies in an effort to increase awareness on all aspects of the technology. Having traced the development, the next section describes the technology.

3.5 CAD SYSTEMS

CAD technology comprises computer software and hardware. In general, these computer components of the system should be made compatible with the users' cognitive processes and physical requirements (Pikaar, 1989). Failure to achieve this will result in some of the problems discussed in Section 3.3.2.

3.5.1 Hardware

For current purposes, the basic CAD hardware can be considered as consisting of three parts: processor, input devices and output devices. The computer processor performs data storage and retrieval, data manipulation and peripheral control; the input devices are used to input design information and system commands; while the output devices display information, including design representations, either on a video display or on a hardcopy medium. The next section will describe in some detail the I/O elements of CAD systems. The aim is to characterise CAD in terms of human factors aspects that are relevant to the thesis, given that CAD is the domain of application for investigating speech-manual integration.

Common configurations of input devices

As described in Chapter 2, several input devices are available for use in carrying out computer tasks. However, they are not as widely incorporated in CAD systems as the tablet and the keyboard. The next section discusses common configurations of input devices in CAD.

Tablet and keyboard

For CAD tasks, the graphics tablet is the most highly used input device for the entry of graphical data from drawings or maps and for the selection of menu items via a screen menu or a menu overlaid on the tablet surface. The tablet exists in two forms, as a large digitising table or as a portable, desktop tablet which occupies a sizeable portion of the desk space. The choice of transducer is largely task dependent. A puck consists of a small frame containing a transparent viewer and incorporating a crosshair cursor. If digitising is the main task, a puck may be the best choice. Alternatively, a stylus can be used for either freehand drawing or pointing. A stylus resembles a pen with a switch activated by pressure on the tip; it may also carry command buttons for controlling basic subroutines.

The alphanumeric keyboard is part of a computer terminal. It is the standard device for input of alphanumeric data and commands into the computer. As mentioned in Chapter 2, the most common type of keyboard layout in use is the QWERTY layout; it is now an official ISO standard. Keyboards may sometimes be extended by function keys and numerical keypads. Function keys are used to simplify the input of commands or to select menu items or move a cursor. If used with a menu interface, selection from menus may be made via the cursor keys on the keyboard.

Mouse and keyboard

The keyboard performs the same functions as described above. The mouse, as discussed in Chapter 2, is a less accurate input device than the tablet for the entry of graphical information. It is, however, best suited to the selection of menu items, and freehand sketching of pictorial data in the form of line drawings. Push buttons may be mounted on the top of the mouse to enable a user to select commands. In the office environment, the mouse has almost become the standard complement to the keyboard, forming an integral part of the popular Windows, Icons, Mice and Pull-down menus (WIMP) interface.

Output devices

There are two kinds of output devices: those which produce hardcopy, and those which generate images/pictures on a visual display terminal (VDT). Hardcopy devices such as plotters, printers and slide recorders are used to output permanent records or copies of CRT and semiconductor displays. These displays are image-producing devices, used for representing graphical and/or textual information.

A CAD workstation usually uses one or two terminals. A terminal consists of a CRT (eg. refresh, raster scan, vector) or non-CRT display (eg. Laser, Plasma Panel, LED), and usually a keyboard. It can be of two types - graphics and alphanumeric. Alphanumeric terminals only display text, while graphics terminals can display graphics and text. Some terminals of each type can provide colour as opposed to monochrome output.

A dual-screen configuration can be of two types: (1) a separate-screen configuration comprising one graphics screen and one alphanumeric screen; or (2) a combined-screen configuration consisting of two graphics screens. In the first type, screen menus are displayed on the graphics screen either in the form of pull-down/pop-up menus, or alternatively, along one edge of the screen. If desired, a prompt area (eg. one line) may be displayed at the bottom of the drawing window, while command line and system messages are displayed on the alphanumeric screen. In the second type, screen menus and prompt lines may be displayed on either of the graphics screens.

In a single-screen system, the screen is used for both graphics and text. In this instance, the prompt response area occupies more space (eg. three lines) at the bottom of the screen. In some systems (eg. AutoCAD), in order to view system messages and/or to enter text, the user will be required to toggle a flip-screen key on the keyboard, back and forth between graphics and text. Due to this inconvenience, single-screen systems are less common in CAD applications, and if used, the screens are usually large to accommodate both graphics and text information.

3.5.2 Software

Given the range of I/O devices from different vendors, ideally, a CAD software package must be able to interface with the different configurations of devices. Thus, most CAD software comes with a configuration program which allows the user to define the hardware devices used in the system. CAD software runs on various operating systems. Operating system software is the interface between CAD application software and the hardware (eg. MS/PC-DOS, OS/2, Unix). The application software comprises programs which perform the particular CAD functions of the system.

2D draughting

2D drawing software includes 2D draughting and other drawing applications which only need 2-dimensional representation. A CAD draughting system is in fact an automated version of the conventional draughting method but it is much more flexible in terms of geometry definition than the manual methods. For example, the basic geometry of draughting: points, lines and circles can be defined in several ways in CAD. Usually, these alternatives are graphically shown on an overlay menu on the graphics tablet or a help menu on the CRT screen so that the user does not have to memorise all the available options. Using 2D draughting has its limitations. The solid object (see solid modelling below) cannot be represented as a solid, hence engineering analysis cannot be performed; the user cannot see the 3D image on a display; and there is difficulty in representing objects which are smoothly curved in two directions (eg. a sphere). To accommodate these facilities, alternative systems have been developed, namely, 3D wireframe, 2.5D drawing and mapping models.

2.5D, mapping and 3D wireframe

A 2.5D model is a 2D model with a constant z axis dimension (eg. a shaft). Mapping models enable digitised terrain data to be processed to produce contour and/or 3D drawings. A 3D wireframe model describes the edges and outlines of curves, thus allowing a part to be modelled as a set of lines or edges in space. Wireframe models are easy to generate and are useful as visual aids. However, since there is no information on the surfaces nor the inside or outside of the object, the notion of solidity is not conveyed. This calls for a 3D solid modelling software.

Solid modelling

Solid modelling is of two types: (1) *constructive solid geometry*, which builds up models in terms of volumetric elements called primitives (eg. blocks, cones, etc.); objects are constructed by the use of set operations (eg. union and difference); and (2) *boundary representation modellers*, which store the product definition in a highly redundant format containing explicit details of all faces, edges and vertices of the object, together with topological information concerning the interconnections between these geometrical entities (Majchrzak et al., 1987). The advantages of solid modelling are: (1) design draughting and analysis can be performed directly on the modelled solids; (2) the same design database can be shared by other design, analysis, etc. modules, thus saving time; and (3) 3D representations of models can be displayed realistically in colour on the CRT screen. Because solid modelling systems are more complex than 3D wireframes, many internal operations have to be performed in order to evaluate the model and to generate the display; thus they run slower and require more powerful computers and larger storage facilities.

Given the above, the CAD system that will be investigated in the thesis is a PC-based system comprising: (1) a dual-screen configuration of separate graphics and text displays; (2) a graphics tablet with a stylus and a QWERTY alphanumeric keyboard; and (3) a 2D draughting software. Details of this system will be described in Chapter 7.

3.5.3 CAD tasks

The tasks that will be presented here are those that are considered relevant to the thesis. This means that other views of CAD tasks (eg. Whitefield, 1986b) are not relevant here, with the exception of Ballay's (1988) view of design as a 'visual task' (see Section 3.2.1).

CAD tasks can be categorised into design and digitising. *Design* includes draughting whereby objects from a plan are reproduced in the form of layout drawings. *Digitising*, on the other hand, involves copying or tracing the objects to generate a drawing. The main differences between draughting and digitising lie in: (1) the type of operations performed, that is, draughting involves more cognitive operations (eg. decision-making), while digitising is more routine; and (2) the range of information utilised, that is, draughting involves a larger set of information inputs (eg. commands, graphical and alphanumeric data), while digitising involves a limited subset of information. As such, the behaviour patterns of performing these tasks are not the same.

To aid designers in selecting devices and techniques for performing graphical interaction (including CAD), Foley, Wallace and Chan (1984) suggest that graphic interaction sequences be decomposed into six fundamental interaction tasks (FITs). These tasks, which are independent of hardware and application domain, form the building blocks from which more complex interaction tasks are assembled. FITs are defined as follows:

- *select* - the user makes a selection from a set of alternatives: the set could be a group of commands or a collection of displayed entities;
- *position* - the user indicates a position on the display, often as part of a command to place an entity at a particular position;
- *orient* - the user orients an entity in 2D or 3D space;
- *path* - the user generates a path, which is a series of positions or orientations created over time;
- *quantify* - the user specifies a value to quantify a measure;
- *text* - the user inputs a text string to annotate a drawing.

Besides the interaction tasks, Foley et al. (1984) also defined four control tasks. The characteristic difference between FIT and control task is that the former specifies, while the latter controls objects already on the display. The four control tasks are defined as:

- *stretch* - the user grasps a feature and moves it to a new position, leaving the remaining features of the object in place;
- *manipulate* - the user causes an object to move in the space by either translation or orientation;
- *shape* - the user causes a smooth curved line or surface to change its general shape according to a positioning device; and
- *sketch* - the user, manipulating a locating device as if it were a pen, creates an object by freehand sketching.

Some of these terms will be used in the thesis to describe CAD task performance. The control task - sketch - however, will not be part of the CAD tasks to be investigated. The next section examines three important ergonomic aspects of CAD.

3.6 HUMAN FACTORS IN CAD

There are various human factors aspects of CAD as described in the CAD literature (eg. Spence, 1976; Majchrzak et al., 1987), such as perception of pattern and function, command dialogue, etc. But the human factors to be discussed here are those of interest to the investigation, and they relate to data entry and information display aspects of CAD.

3.6.1 Commands

The choice of command names and design of command dialogue are important concerns in interface design (eg. Barnard, Hammond, Morton, Long & Clark, 1981; Carroll, 1982; Scapin, 1981). Some of the issues usually mentioned in connection with choice of command names are: (1) ease of learning and memorability; (2) task specificity and naturalness; (3) consistency among names; and (4) size of command set (eg. Nickerson, 1986; Majchrzak et al., 1987).

Some command names are chosen as metaphors of non-computerised tasks (eg. cut and paste). For CAD, many of the tasks which are used in manual draughting (eg. sketch, chamfer) are employed by system developers as a metaphor or analogy for the computerised task. This should make command names easy to remember, which in turn might help the optimisation of transfer between systems, especially for a drawing board user who becomes a CAD operator. Another issue is whether to choose command names which are specific to the task or natural to the operator. Studies (eg. Dumais & Landauer, 1981) have shown that novices initially chose names that were non-specific but found these to be misleading as they gained more experience. But specificity of the command name should not be compromised for naturalness.

The choice of names is also an issue in the design of menus. Dumais and Landauer (1981) pointed out that success at using systems tended to be high with systems having relatively small command sets (usually less than 100 items) that are well partitioned into non-overlapping categories. CAD tasks are slightly different from other computer tasks (eg. word processing) in that the operator must enter more commands to perform the task (see Section 3.5.3). The organisation of commands in a menu is therefore crucial in order to aid visual search. Some generic CAD commands and their functions will be discussed in Chapter 7, taking AutoCAD as an example.

There are different ways of entering commands into the computer (see Chapter 2). The following methods will be investigated in this thesis. In CAD, commands can be input via: (1) alphanumeric keys on the keyboard, by simply typing in the command name; (2) function keys (on the keyboard, puck or stylus), by depressing the designated-command key; and (3) menus - screen menu or tablet menu, by positioning the transducer over the desired menu item and depressing the 'pick' button, or by pressing the menu cursor key on the keyboard. The use of function keys is less general and less flexible than menu selection.

3.6.2 Graphical and alphanumeric data

The entry of a command is usually supplemented with additional information (eg. parameters) for executing the command function. Two types of information will be considered here, namely: (1) graphical data or screen coordinates; and (2) alphanumeric data.

The coordinate system in CAD is based on the Cartesian (or world) coordinate system, in which points are addressed by their x and y coordinates. In 2.5D and 3D systems, a third dimension (z axis) has to be specified. One of the main problems with creating and editing drawing objects is the accurate specification of a position in the drawing where the object entity (eg. line, circle, arc) is or should appear. This may be defined as: an absolute position measured as a displacement from some fixed position (usually the bottom left corner of the drawing window) or as a relative position, expressed as a displacement from the previous point (Newman & Sproull, 1979).

The problem becomes more complex with 3D systems when trying to line up the objects together in a way desired by the user. Recent systems have attempted to resolve this issue by providing two types of coordinate system: one 'world' and many different 'user' coordinate systems. The world coordinate is fixed in terms of x, y and z axes; the user coordinate can be defined by the user anywhere and at any angle in 3D space (Bright, 1988). These user-defined coordinates can then be saved with the drawing.

The creation of entities sometimes requires a specification of parameters, usually a numeric value that specifies height, width, length or distance. This type of information is termed here *numerical data*. In addition, drawing objects sometimes require some form of annotation or labelling which is made by inputting alphanumeric characters, known here as *textual data*. The design of alphanumeric characters is a central issue in information display, particularly those concerning legibility and quality of the character. Factors such as size, shape, colour, orientation, shade and texture are important ergonomic considerations in character recognition (eg. Newman & Sproull, 1979; Davis & Swezey, 1983).

In a dual-screen configuration with separate screens, deciding the display size of characters is an important concern. This is because the text appears on one screen while the graphics on another. Therefore, mapping of text and object size can be a problem, especially to novices. Character sizes are conveyed in terms of visual angle sizes. Characters larger than 25 minutes of arc are considered inappropriate because they disrupt reading time due to excessive eye movements (Majchrzak et al., 1987). In other words, the layout of the information on the screen and the graphical representation of the objects can affect the interaction process.

3.6.3 System feedback

Some studies have shown that task performance depends on the presence and the quality of feedback (eg. Shneiderman, 1982). A CAD system provides one or two standard forms of feedback. One of the most basic is the feedback of a cursor that follows the coordinate input device in the form of a crosshair. Therefore, the choice of an input device is partly determined by its ability to provide this visual feedback to the user. An analogous form of feedback is the echoing of typed characters.

The types of feedback available in CAD systems are classified into three types (Newman and Sproull, 1979):

(1) feedback from the command interpreting process, informing the user whether the command has been accepted, what stage of executing the command has reached, and whether an error condition has arisen;

(2) feedback from the application database, mainly for item selection feedback (eg. selected item is highlighted); and

(3) feedback unrelated to command interpretation or to the database, eg. cursor feedback, character echoing, etc.

Command feedback, that is, feedback from the process that interprets the user's commands is needed for several reasons:

(1) to show the general effect of the user's next command (eg. which menu item has been selected);

(2) to indicate if the command is erroneous in some respect (eg. it cannot be applied to the selected operand). For this, the user needs an immediate error response. Any delay will make it difficult to recreate the mental context that is needed to restate the command;

(3) to confirm, if the execution of the command is very slow, that the computer is still working on the user's request; and

(4) to help the user with the next command entry in the form of prompts. The user responds to the prompts or messages by entering answers via the input devices.

In short, the role of feedback is crucial to effective interaction with the computer.

3.6.4 Conclusion

The above describes some human factors aspects of CAD that are relevant to this research. They concern: (1) the choice and design of command names; (2) the layout and display of graphical and alphanumeric data; and (3) the availability of feedback to aid task performance. This information will be particularly useful in attempting to understand the distribution of attention between different parts of the CAD displays. Having described some underlying issues in CAD, the next section reviews some studies that have addressed human factors issues in CAD.

3.7 EMPIRICAL STUDIES ON HUMAN FACTORS IN CAD

The purpose of this review is to highlight issues that might have some bearings on the problem investigated in the thesis. As such, the review is not intended to be all-encompassing. To guide the selection process, the following criteria are used.

(1) the emphasis in the research is on CAD tasks and/or CAD representations;

(2) the research aims to provide an understanding of performance and/or behaviour issues, using modelling techniques;

(3) the research is based on empirical data (gathered via experimentation, direct observation, or interview methods), or development of a methodology.

Within this set of criteria, the following five studies are selected.

Whitefield (1986b; 1986c) compared two groups of mechanical engineering designers (n=4), one using a CAD system and one using drawing boards. Their design activity was recorded on videotape, using concurrent verbal protocols as the primary source of data. The purpose of the comparison was to describe the design knowledge used by each group in terms of what classes of knowledge they apply and how they apply it. Using a blackboard framework (derived from the HEARSAY-II speech understanding system) to construct the model, the analysis provides an understanding of differences in performance between computer aided and unaided design activity.

The findings showed that CAD designers recruited more drawing knowledge than drawing board designers in designing a television casing. In particular, the findings contradict the claim that CAD "...takes over the dull parts of designing and leaves the designer free to concentrate on the creative and interesting parts. Rather than unburdened, the designer is further hampered by the demands of CAD system operation, because producing a drawing using CAD is more complex and demanding task" (Whitefield, 1986c, p. 96). This finding suggests that using CAD requires additional drawing skills to manipulate the computer objects on screen with the input devices. On the basis of this, Whitefield (1986b) commented that the development of CAD has not depended on studies of design.

Two important limitations of the study are: (1) the use of verbal protocol technique which in itself has certain requirements (Akin, 1984); and (2) the use of a small number of subjects; as such it was not possible to test statistically the differences in the way knowledge was recruited by both designer groups.

Antin (1988) compared experienced and novice users (n=12) of an interactive 3D CAD system on objective performance (completion time, errors) and subjective preference measures. The aim was to evaluate the usefulness of menu selection for both groups of users, and to determine the effects of information presentation aspects of a menu and its input mode on menu use. Three input modes were compared: menu selection, command entry and a combination of the two modes. Although command entry produced performance that was superior to the other two modes, there was a strong user preference for the combined mode. The notion that menus are viewed as a hindrance by experienced users was not supported by the findings.

Card et al. (1983) analysed the performance of a CAD VLSI circuit task in the context of the GOMS model (see Chapter 1). The purpose of the model is to predict performance times so that the information could be used: (1) in the training of CAD operators; (2) for identifying levels of expertise between novices and experts; and (3) for design of CAD systems. An expert CAD operator modified an existing design for about 40 minutes. By recording his behaviour and verbal protocols on videotapes, the analysis provided a means of determining the overall goal structure, and decomposing it into subgoals, unit tasks and events. A major limitation of the study was the use of a single expert (with a year's CAD experience) to provide accurate time estimates of the tasks and using this as a basis for generalizing to other operator behaviour.

Sharit and Cuomo (1988) described a cognitively-based methodology for evaluating human performance on CAD tasks. The CAD task domain to which the methodology is applied is architectural design; it could be used to evaluate the utility of various CAD systems that are being developed to aid architects. The purpose is to understand how novel CAD systems affect design performance for: (1) different types of tasks; and (2) different levels of complexity for a particular task. Task complexity is seen as a function of three attributes: criteria, entities and relationships. Applying data summarization and modelling techniques to error measures enables human performance to be explained as a function of various cognitive demands (perceptual, memory, decision making and motor loads). The implications of these demands for CAD tasks were made, but in the absence of empirical data to support the validity of the claims, the effectiveness of the methodology is yet to be proven.

The last study considered is by Dowell (1986). The aim of the research was to evaluate the usability of different forms of CAD graphical representations for the mechanical engineering design task. The research identified three generic forms in which mechanical objects are represented, namely, the 2D orthographic, 2.5D wireframe and the 3D solid model. Some design engineers were given the task of assessing a carburettor design against the listing of its required functions. Each designer was observed and verbal protocols were recorded during the performance of the task. A qualitative analysis of the data revealed some differences between the groups in design fault identification on the basis of the time taken to complete the task (resource cost) and errors committed (cognitive effort). An implication of the study is that the type of graphical representation has an effect on CAD performance. In short, the findings suggest that a CAD characteristic can incur performance costs to the user.

In conclusion, the above studies demonstrate that the use of CAD system can affect: (1) performance in terms of knowledge recruited, resource cost and cognitive effort expended to perform the task; and (2) behaviour, resulting in non-optimal deployment of resources, whereby more resources are used to manipulate the input devices than to perform mental operations. The method of combining verbal protocols and behaviour protocols to study the design process might prove effective in understanding interactive behaviour. A common methodological flaw is the small sample size and the lack of statistical analysis to support the findings. The studies have contributed to the development of user models as design tools for system designers. On the basis of these findings, it could be inferred that the use of CAD systems might produce non-optimal behaviour, which in turn may incur performance costs, the problem addressed by this thesis.

3.8 SUMMARY

This chapter provides a general introduction to the technology of CAD and is intended to give readers an overall concept of what CAD is and what it can do. Previous work that examines the human factors aspects of CAD is described to ascertain what has been done and what needs to be known for this research. Configuration of a demonstrator CAD system will take into account the human factors issues raised here, specifically, commands, data and feedback, besides the hardware and software aspects. The specific CAD system will be described in Chapter 7.

NEXT CHAPTER HIGHLIGHTS

In Chapter 1, the problem of using manual input devices was identified as resulting in non-optimal behaviour, which in turn may incur performance costs to the user. The role of Chapter 4 is to suggest a methodology as the basis for an empirical evaluation of CAD systems. The methodology employs behaviour protocols and performance indices as sources of data for understanding the effects of systems.

CHAPTER 4

A Behaviour-based Methodology for Empirical Evaluation of CAD Systems

Overview

4.1 Introduction

4.1.1 Identifying features of the methodology

4.2 System assessment

4.2.1 System specification

4.3 Determining optimal design behaviour: CAD versus traditional draughting

4.4 Behaviour protocol analysis in design

4.4.1 Protocol analysis

4.4.2 The use of video data in protocol analysis

4.4.3 Automated scoring of video data

4.4.4 VITAS

4.5 Behaviour and performance measurement

4.5.1 Behaviour concepts

4.5.2 Performance concepts

4.5.3 Behaviour-performance distinction

4.6 System behaviour modelling

4.7 Human factors guidelines

4.8 Summary

Next chapter highlights

CHAPTER 4

A Behaviour-based Methodology for Empirical Evaluation of CAD Systems

OVERVIEW

As identified in Chapter 1, the unitary use of manual input devices for performing CAD tasks may result in sub-optimal behaviour for the user. CAD as a domain for exploring the problems was described in Chapter 3, while Chapter 2 discussed input devices and their human factors. To investigate further the problem, a methodology for evaluating CAD systems is described in this chapter. The methodology is based on a framework that describes the I/O processes presumed to underlie performance in design activities, using behaviour protocols and performance indices as data.

The methodology is intended to be system and technology independent, and acknowledges that current and future developments in CAD systems are likely to change the way in which the designer interacts with the computer. In describing the methodology, an attempt is made to define terms precisely. The clarity is needed: (1) to avoid confusion between the terms as used here and elsewhere; (2) to enable operationalisation of the concepts in conducting the research; and (3) to render better interpretation of the findings.

4.1 INTRODUCTION

The features of this methodology meet the requirements for achieving the research goals outlined in Chapter 1. These goals refer specifically to the need for:

- understanding better the capabilities and limitations of the user in relation to the input devices employed to perform the task, based on a characterisation of the I/O processes of an interaction; and
- developing data analysis techniques, including a modelling method for understanding system behaviour, a computerised technique for behaviour protocol analysis as an efficient means of quantifying system behaviour, and a statistical tool for analysing the possible effects of types of systems on behaviour and/or performance.

An overview of the research framework outlining the methodology is depicted in Figure 4.1.

4.1.1 Identifying features of methodology

This section presents an overview of the methodology. Details of each feature in the methodology will be dealt with in the rest of the chapter. With reference to Figure 4.1, the main

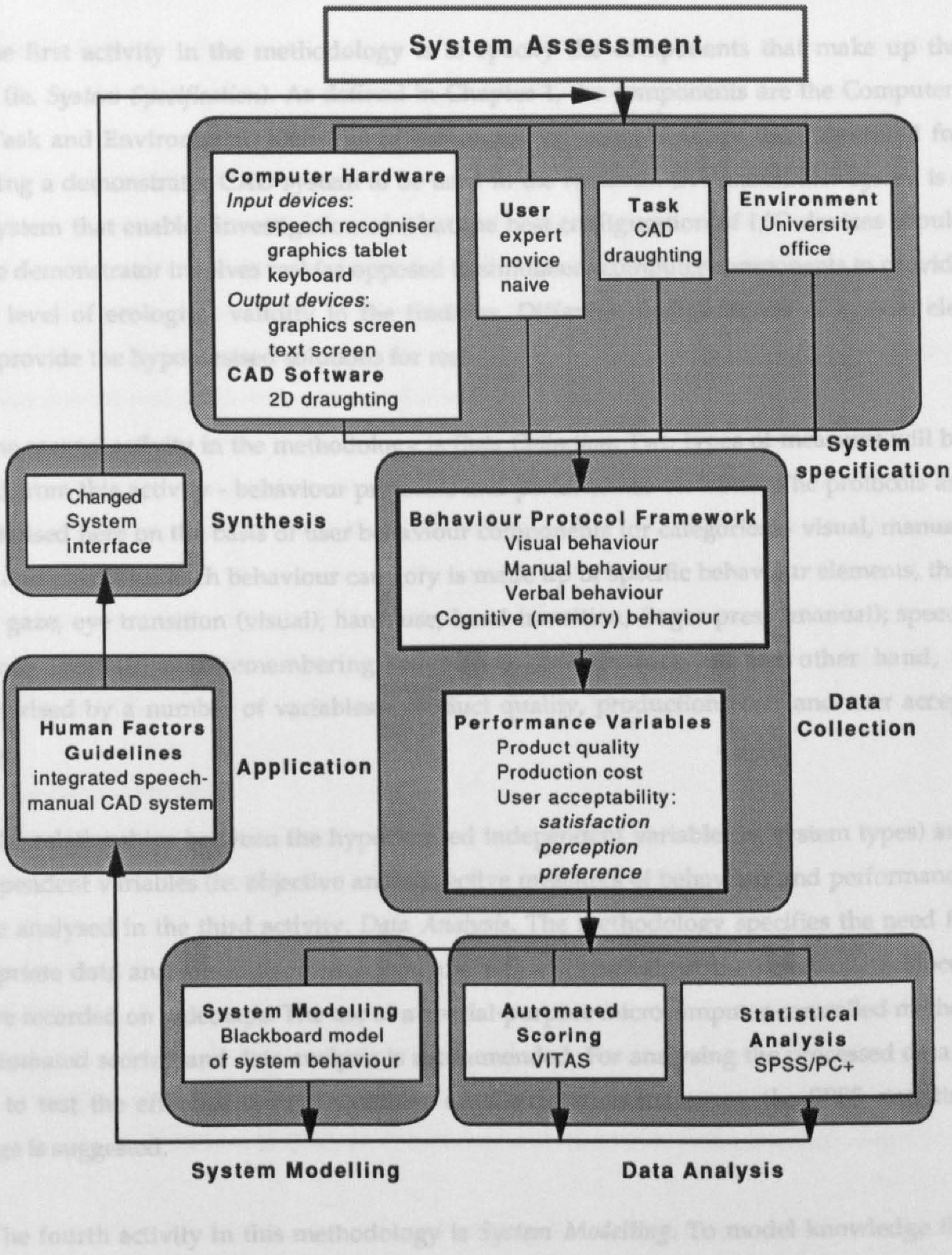


Figure 4.1. Overview of a behaviour-based methodology for evaluating CAD systems

features of the methodology will be described, taking in sequence the activities labelled in the shaded boxes. These activities occur in different phases of the research process, outlined in Chapter 1, namely, analysis, generalisation, particularisation and synthesis.

The first activity in the methodology is to specify the components that make up the system (ie. *System Specification*). As defined in Chapter 1, the components are the Computer, User, Task and Environment. Elements of the computer component are then identified for specifying a demonstrator CAD system to be used in the research. (A *demonstrator system* is a CAD system that enables investigation of what the best configuration of I/O devices should be.) The demonstrator involves real (as opposed to simulated) computer components to provide a high level of ecological validity to the findings. Different configurations of system elements provide the hypothesised solutions for test.

The second activity in the methodology is *Data Collection*. Two types of measures will be derived from this activity - behaviour protocols and performance variables. The protocols are characterised here on the basis of user behaviour components (or categories) - visual, manual, verbal and cognitive. Each behaviour category is made up of specific behaviour elements, that is: eye gaze, eye transition (visual); hand use, hand transition, finger press (manual); speech utterance (verbal); and remembering (cognitive). Performance, on the other hand, is characterised by a number of variables - product quality, production costs and user acceptability.

The relationships between the hypothesised independent variable (ie. system types) and the dependent variables (ie. objective and subjective measures of behaviour and performance) will be analysed in the third activity, *Data Analysis*. The methodology specifies the need for appropriate data analysis tools for accurate and efficient analysis of the behaviour protocols that are recorded on videotape. The use of a special-purpose microcomputer-controlled method for automated scoring and data analysis is recommended. For analysing the processed data in order to test the effect of system solutions on the criterion measures, the SPSS statistical package is suggested.

The fourth activity in this methodology is *System Modelling*. To model knowledge that is recruited during CAD performance, a theoretical framework from the science base is used. The framework is based on a blackboard model of design. Data from the analyses are then used to construct the model. The model serves as an analytical tool to relate the problem to the solutions and as a basis for the guidelines.

To complete the research process, the methodology describes the application of the research findings to generate some human factors guidelines. This *Application* activity is a particularisation of the science support representation. The final activity is *Synthesis* which involves validating the suitability of the guidelines in the workplace.

This concludes the very brief description of the research process. Much of the remainder of this chapter will be directed at operationalising the concepts to be used in the thesis.

4.2 SYSTEM ASSESSMENT

The term *assessment* will be used interchangeably with the term *evaluation* to refer to the process in which relationships between the underlying variables are observed and explained, and some judgement is made about the system. The process spans a cycle of research activities from an initial problem specification in the real world to an investigation of solutions in a simulated context and synthesising the output in the workplace.

In this research, then, *system assessment* involves a series of empirical studies (observational and experimental) in which the different types of system (independent variable) are systematically introduced to ascertain whether and how they affect user behaviour and task performance (dependent variables). The purposes of this assessment are: (1) to identify non-optimal target behaviours; (2) to suggest solutions to reduce the non-optimal target behaviours; and (3) to evaluate the outcome of the solution 'treatment'.

4.2.1 System specification

As mentioned above, and in Chapter 1, the system components are Computer (including hardware and software peripherals), User, Task and Environment.

The *computer* hardware of interest in this research includes both input and output devices of dual-screen CAD systems. The input devices are speech recogniser, graphics tablet and keyboard, while the output devices are graphics and text screens. The computer software used is a 2D draughting package for the IBM-PC range. The demonstrator system, including details of the software, is described in Chapter 7, together with factors determining their selection.

The *users* employed in the research include *experts* (defined as those with more than a year's CAD experience plus drawing board experience); *novices* (with minimal CAD and/or drawing board experiences plus computer experience); and *naive users* (with no CAD experience and little or no computer experience). Data from expert users provide the basis for: (1) specifying the initial problem concerning the use of unitary manual input; and (2) characterising user performance as a function of skill level. Novices and naive users, on the other hand, provide information for identifying user difficulties in learning and using the demonstrator system. Given that the CAD software investigated is no longer only a tool for experts (eg. designers, CAD operators, etc.), but also for occasional users (eg. designer trainees, architectural students, etc.), the information is relevant to system development. Since different users are known to have different needs of the system (eg. Allwood, 1986; Antin, 1988), characterisation as a function of user type helps to identify proficiency level and training needs.

The *task* consists of CAD tasks, in particular, draughting tasks whereby objects from a plan are reproduced on screen in the form of perspective (layout) drawings. This contrasts with: (1) the complete design task which involves more than just draughting; and (2) digitising, which involves copying or tracing of objects to generate a drawing (see Chapter 3). A comparison between design behaviour that is expected with CAD use and with traditional draughting is made in Section 4.3. The choice of draughting tasks is based on the fact that it involves more CAD operations, thereby permitting better assessment of device utility. Other task aids that seem necessary to support CAD performance include drawing plans, a CAD reference manual, speech vocabulary lists, calculators, writing tools, etc. These are called here task-related tools.

The *environment* is a controlled laboratory setting, that of an office. The choice of an office environment is because CAD activity in general takes place in offices - either design offices (if design experts) or university offices (if design students). Conducting the research in an office environment, in addition to using real systems (computer I/O, task, people), provides ecological validity to the research.

The need for ecological validity and the use of the 'real world' as the laboratory for human factors research is emphasised by Chapanis (1967). This leads on to the next section, the identification of representative CAD tasks for investigation.

4.3 DETERMINING OPTIMAL DESIGN BEHAVIOUR: CAD VERSUS TRADITIONAL DRAUGHTING

Design involves the creation of specifications for constructing objects that satisfy particular requirements (see Chapter 3). Draughting, as stated above, is a subset of design tasks. In traditional design, draughting takes place on the drawing board using tools such as pencils, slide rule, compass, set square, etc. Calculators have supplanted slide rules, thus aiding the designer in numerical calculations and analysis. With the introduction of computers to support the design process, the functions of the drawing board are 'transferred' to the graphics screen. The equivalents of the input tools (pencils, etc.) are input devices, usually the graphics tablet with a transducer, either a puck or stylus.

A significant difference between manual design and CAD lies in the entry of commands to operate the system, in addition to the input of drawing data. The consequence is that CAD designers have an additional task of manipulating computer objects on screen, besides making decisions and solving problems in the design domain space. In this respect, the CAD draughting task differs from the drawing board draughting task, and the way in which the designer interacts with the system tools also changes.

Given the above, it is necessary to define what effective (optimal) draughting behaviour is. The aim is to provide criteria for evaluating whether the types of system being investigated here produce the expected (or normative) behaviour (see Chapter 5), that is equivalent to traditional draughting behaviour. This requires a distinction between what are considered *primary* and *secondary* task activities in CAD performance. Drawing, which includes the entry of information (commands, graphical and alphanumeric data), and the navigation of system tools, is treated here as a primary activity. Secondary, are activities such as handling the plan, system manual, and other task tools. If drawing is the main activity, then, a greater proportion of the resources should be expended to undertake this activity.

Therefore, drawing a parallel between traditional draughting and CAD draughting tasks, it could be said that:

In traditional draughting,

If	the primary task is drawing
Then	eyes should be on the drawing board for greater proportion of the time on task (relative to other targets/elsewhere); and hand should manipulate drawing tools (relative to non-task tools).

In CAD,

If	the primary task is drawing
Then	eyes should be on the graphics screen for greater proportion of the time on task (relative to other computer/task tools); and hand should manipulate the graphics tablet for drawing (relative to other computer/task tools).

The above defines what *optimal* draughting behaviour is. Extending this to the research, it becomes possible to identify specific types of optimal behaviour *vis-a-vis* device use, as defined in Section 4.5.1.1. An analysis of behaviour protocols would provide the means for distinguishing the distribution of resources in CAD performance.

4.4 BEHAVIOUR PROTOCOL ANALYSIS IN DESIGN

In this section, the use of protocol analysis as an investigative technique will be discussed, together with procedures for analysing the protocols.

4.4.1 Protocol analysis

Protocol analysis is a technique devised to infer the information processing mechanisms underlying human problem solving behaviour (Newell, 1968). A *protocol* is the recorded behaviour of the problem solver. It is usually in the form of traces or recordings of the overt behaviours, such as notes, sketches, video or audio recordings of behaviours, etc. Verbal protocols are a subset of protocols, obtained through verbalization or 'thinking aloud' procedures (eg. Whitefield, 1986b; Akin, 1984; Schoenfeld, 1983). Hence, the verbal data are assumed to reflect the thought processes of the problem solver. Because verbal data analysis is complex and time-consuming, the sample size is usually small.

This study, however, is based on overt perceptual-motor actions, subsequently termed *behaviour protocols*. The protocols are recorded on videotapes for analysis. (The term *videotape* as used here refers to the use of standard videocassettes.) This procedure enables the record of a reasonable number of subjects (n >6), and the analysis of each subject's protocol for a longer period of time in order to obtain sufficient observations of desired behaviour types. Hence, the problem of sample size is reduced. The conclusions reached at the end are generalisations about the consistencies between many users of the system rather than within each user.

The use of behaviour protocols is consistent with psychological studies that examine human and animal behaviour in various social, clinical and ecological settings (eg. Summerfield, 1983; Slater, 1978). They provide a rich source of information for understanding the interaction between user and computer.

4.4.2 The use of video data in protocol analysis

As stated above, the protocols are video records of behaviour. Video is a powerful medium for capturing and conveying information about how people interact with each other and with machines, specifically computers (Mackay, Guindon, Mantel, Suchman & Tatar, 1988; Laws, 1988). It provides a permanent record of sequential streams of natural observations, some of which are subtle (gestures and eye movements), and difficult to capture in any other form. Video can also preserve the context as well as the content of a session and provide multi-faceted, qualitative data that can be analysed in a number of different ways (Dowrick & Biggs, 1983).

Within HCI, video has been used by researchers to study human-human interaction (eg. cooperative work, teleconferencing) and human-computer interaction (eg. product testing, user interface development). In the latter instance, the data could be used to identify problems with hardware and software products, and to provide feedback to system developers of the required changes to the interface. Given the potential of video use in HCI, why is its use limited?

There are several possible explanations: (1) the lack of familiarity and lack of understanding of its benefits; (2) video's reputation for being cumbersome to edit; (3) the difficulty associated with publishing video data; (4) there are no equivalent statistical methods that could provide verifiable summaries of video-based results; (5) the richness of the data which makes it difficult to compress in meaningful ways; and (6) the process of analysing video material which can be very time-consuming (Mackay et al., 1988). However, researchers have begun developing computer-assisted techniques for capturing and analysing video data that attempt to address these problems.

The various constraints and drawbacks to video analysis, especially those concerning 'filming' issues, reliability and validity of measurements will be discussed in Chapter 12, as part of the problems experienced in the investigation. The available literature on video technology (eg. Dowrick & Biggs, 1983; Berger, 1978) provides comprehensive coverage of the technical issues that must be considered in any video work. The next section examines how automated scoring is done within this research.

4.4.3 Automated scoring of video data

The advent of computer interactive video technology provides a means of controlling the functions of the videotape recorder and player via the computer. This is achieved by recording an electronic time code signal (supplied by the computer) onto an audio channel of the videotape, thus providing frame-accurate indexing of the video material. The time code is then the key to efficient indexing and electronic control of videotape. Programs are then responsible for controlling such features as: searching the tape for a point in time (fast forward and rewind modes), playing a particular section of videotape, and editing selected frames.

In recent years, some prototypes of video interactive systems for computerised analysis have been developed, such as Video Interactive Technique for Automated Scoring (VITAS) (see Laws, Summerfield, Watson & Elton, 1986), Computer Assisted Sports Evaluation (CASE) (see Franks & Nagelkerke, 1988) and Protocolling and Retrieval of Audio-Visual Data Analysis (PRAVDA) (see Clarke & Ellgring, 1983). This study employs the VITAS system in the scoring and analysis of videotaped behaviour protocols.

4.4.4 VITAS

The first generation VITAS, developed jointly by Birkbeck College, University of London, and the United Medical & Dental Schools of Guy's and St. Thomas's Hospitals, was originally designed for clinical psychology purposes. The second generation VITAS is currently being developed at STC Technology Ltd., UK, using a more powerful computer system (ie. SUN workstation) than the Apple IIe microcomputer used in the earlier VITAS.

This study, however, uses the first version of VITAS (see Laws et al., 1986 for details). The system is composed of an Apple IIe microcomputer, linked to a JVC U-matic videocassette recorder (VCR) via a specially developed interface (ie. a circuit board). A standard monitor, linked both to the VCR and the computer, is used to display the video recordings. Figure 4.2. shows the configuration of the system. The key element in controlling the operation of the VCR, as mentioned earlier, is the electronic time code that is stamped on the tape by the computer. These codes are in fact a series of sound pulses that are laid down on audio track 1 of the videotape. This time code provides the time base for the scoring procedure so that each behaviour scored is stored in association with its frame code. These data are stored on disk and may be retrieved for data processing purposes.

Two scoring issues require mention here. First, the level of detail required for behaviour scoring must be defined so that the data set produced addresses the problem under investigation. In this study, the problem-solution specification will determine the level of detail needed. Second, the criteria for scoring each behaviour must be clearly specified. In this study, description of target behaviours and the metrics (frequency, duration) to be derived from these behaviours are defined prior to the investigation. Section 4.5. gives the details of the categories.

The computer performs a number of functions, two of which are: (1) it drives the VCR so that a predetermined frame may be accessed and displayed; and (2) it reads the current frame code corresponding to the moment when an event of interest occurs. VITAS programs, written in BASIC, are responsible for editing time pulses, playing video scenes, stopping the tape, etc. Behaviours defined for a particular analysis are presented on a menu, and the scorer selects the category that relates to that frame. At the end of the scoring session, all time and behavioural codes are stored on computer disk for later retrieval and analysis. An example of the menu used in scoring types of visual behaviour and the derived listing of action/time codes is shown in Table 4.1.

Laws (1988) estimated that using visual time code as the time base, manual scoring can take between 8 and 10 times real time. Using automated scoring, this ratio is reduced to

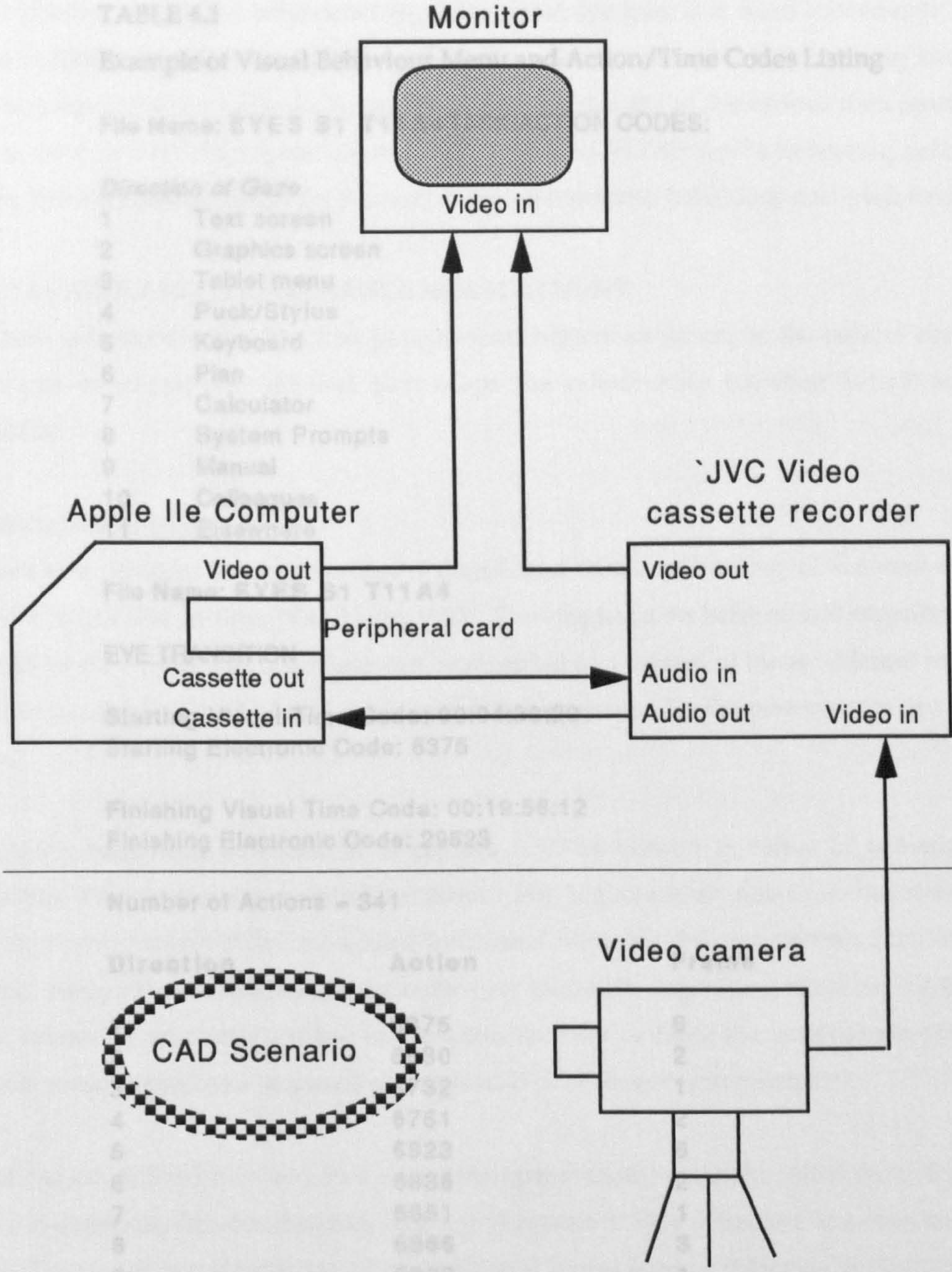


Figure 4.2. VITAS configuration showing the hardware link

TABLE 4.1

Example of Visual Behaviour Menu and Action/Time Codes Listing

File Name: **EYES S1 T11A4**TASK/ACTION CODES:

Direction of Gaze

1	Text screen
2	Graphics screen
3	Tablet menu
4	Puck/Stylus
5	Keyboard
6	Plan
7	Calculator
8	System Prompts
9	Manual
10	Colleagues
11	Elsewhere

File Name: **EYES S1 T11A4**

EYE TRANSITION

Starting Visual Time Code: 00:04:30:29
Starting Electronic Code: 6375

Finishing Visual Time Code: 00:19:56:12
Finishing Electronic Code: 29523

Number of Actions = 341

Direction	**Action**	**Frame**
1	6375	6
2	6630	2
3	6732	1
4	6751	2
5	6823	6
6	6836	2
7	6851	1
8	6866	3
9	6903	1
10	6967	2
11	7154	3
12	7185	2
13	7301	3
14	7339	2
15	7414	3
:	:	:
:	:	:
341	29523	2

between 2 and 4 times real time, depending upon the complexity of the analysis being performed. In addition, programs can be written to calculate the frequency and duration of actions scored, including coincident behaviours (eg. concurrent eye gaze and hand movement). Thus, using an automated system can help to reduce time and effort in both the scoring and data processing stages of video analysis. Appendix 1 gives a summary of the various data processing programs used in VITAS analysis. Having described how VITAS works in scoring behaviour protocols, the next section defines the concepts that characterise behaviour and performance.

4.5 BEHAVIOUR AND PERFORMANCE MEASUREMENT

This section is in three parts. The first part presents behaviour concepts; the second describes performance concepts; and the last part maps the relationship between behaviour and performance.

4.5.1 Behaviour concepts

Behaviour is a process, its organisation is manifested in the patterning of movements and postures in space and in time (Van Hoof, 1982). The emphasis on behavioural organisation is influenced by the notion that behaviour can be regarded as a system of many different routines. Each routine consists of a set of acts, coordinated as instruments for the performance of a certain function.

Treating behaviour as hierarchical enables it to be viewed in terms of activities and subactivities. That is, one can partition behaviour into higher-order molar functional units (or *gross actions*), and lower-order molecular functional units (or *fine movements*). Each of these functional subsystems mobilises certain behaviour elements, organising them on the basis of internal rules that take into account information fed into or from the external situation. So, behaviour presents itself as a sequence of actions and subactions (or movements).

Actions are defined here as coordinated perceptual-motor activities, often requiring some mental computation, decision-making, and use of memory. This definition of action takes into account that actions are dependent on memory and motor control (Norman & Shallice, 1981; Kerr, 1983). A gross action (eg. a hand reaching for a distant object) involves contractions and use of the large muscles of the body (eg. arms and shoulders). A fine action (eg. a finger pressing a stylus button) does not involve the large muscles. This distinction between gross and fine behavioural actions enables user behaviour to be categorised into:

(1) **gross actions** such as eye gaze/transition, hand manipulation/transition, speech production, etc.; and

(2) **fine movements** such as finger press, eye saccade, vocal chord vibration, etc.

This distinction will be used in the analysis of CAD system behaviour.

Behaviour, as defined in Chapter 1, expresses the means by which the system accomplishes its task. Since many different behaviours can produce the same performance, system behaviour is orthogonal to system performance. System behaviour can be expressed in terms of individual system component behaviours - user behaviour and computer behaviour. These behaviours are subject to limits within which they operate in performing tasks. For the user, the behavioural limits relate to capacities such as memory capacity, attention span, manual dexterity, etc. For the computer, behavioural limits relate to capacities such as processing capability and physical configurability.

The modes for expressing behaviour are termed *input* and *output* (I/O). User input modes are visual and auditory, represented by eyes and ears, respectively. User output modes are manual and haptic, represented by hands, fingers and feet; and vocal/verbal, represented by voice/speech. Like the user, computers communicate via physical media termed input and output (see Chapter 2). In user-computer communication, the observable form of behaviour is explicit perceptual-motor actions by the user (eg. eye or hand transition), and explicit computer outputs (eg. changes in display state; screen(s) displaying prompts or system messages).

The relationship between user and computer I/O is an inverse one, such that the user's input mode maps on to the computer's output mode (ie. visual input : screen output), while the user's output mode maps on to the computer's input mode (ie. manual output : graphics tablet/keyboard input; verbal output : speech recogniser input). In this regard, the term I/O will be used here to refer specifically to the various physical computer devices (linking these with the respective user I/O modes). So, the term *speech input* means input device (speech recogniser) linked to user output mode (speech or verbal). Similarly, the term *manual input* means input devices (graphics tablet and keyboard) linked to user output mode (manual).

In Chapter 1, evidence (eg. Van der Heiden & Grandjean, 1984, etc.) was presented to indicate that spending time gazing away from the screen(s), and transiting between input devices, visually and/or manually, may incur performance costs. In terms of visual behaviour, it tends to reduce visual attention on primary drawing activity (see Section 4.3). In terms of manual behaviour, it keeps the hand(s) busy for a substantial proportion of the time on task. The same could be said of verbal repeats of speech utterance via a recogniser. Ideally, any verbal input should be verbalised once and recognised. Repeating the same word again, sometimes several times in order to be recognised, increases time on task (Sperandio, 1987) and may incur some costs (eg. fatigue) to the speaker.

The above behaviours are considered non-optimal for the reasons given above and in

Chapter 1. It therefore becomes necessary to define what constitutes optimal system behaviours in this research in order to assess the extent to which the CAD systems to be used are effective. A CAD system is considered effective if:

(1) eye gaze to graphics screen is significantly greater than eye gaze to other computer and task tools (text screen, graphics tablet, keyboard, plan, manual, speech list, etc.). The more frequent and the longer the duration of eye gaze to the graphics screen, the better able is the user to perform the draughting task. This behaviour is therefore optimal, leading to enhanced performance.

(2) hand manipulation of graphics tablet and keyboard is low. Therefore, the less frequently and the less time the hand is used for operating these devices, the better is the user's performance. This is because in general the hands are less busy, enabling the user to allocate fewer attentional resources to device use. Thus, there will be fewer eye and hand transitions. Therefore, low use or manipulation of the input devices is considered optimal in terms of behaviour and performance.

(3) verbal repeat of speech input due to confusability aspects of speech is low. Less frequent repetition of speech utterances due to recogniser errors incurs less production costs, thus facilitating more efficient performance. Therefore, high single pass recognition and low repeat of commands are optimal behaviours, resulting in better performance.

(4) verbal repeat of commands due to forgetting errors is low. The less frequent the verbal repeat caused by failure to remember which input device is allocated to which data type, the lower the production costs. This means the system places less demand on the user's memory. Therefore, low verbal repeat due to forgetting is considered optimal in both behaviour and performance terms.

Behaviour is quantified here in terms of frequency and duration. *Frequency* refers to the number of behaviour occurrences per unit time (expressed in secs-1). *Duration* refers to the percentage of total time spent on each behaviour type (ie. relative duration). Measuring the time spent per behaviour type enables a comparison of the different costs incurred by the system in performing the task.

Given the above, **optimal behaviour** is expressed as:

$$Bc = fd_n$$

where

Bc	is the behaviour category (eg. visual, manual, etc.)
fd	is the frequency and/or duration of behaviour type
n	is the behaviour type related to the behaviour category (eg. graphics screen gaze for visual behaviour, etc.)

To distinguish between the different types of eye transitions within visual behaviour, the following terms will be used. Eye transitions from graphics screen to text screen, and back, will be termed *between-screen* transitions, while those from the tablet to the keyboard, and back, as *between-device* transitions. Transitions from the screens to input devices will be called *off-screen* transitions, while those from elsewhere (input devices, task tools) to the screens as *on-screen* transitions. The term *within-screen* transitions refers to transitions between points within the graphics or text screens.

4.5.2 Performance concepts

Performance, as defined in Chapter 1, expresses the effectiveness of the system in accomplishing tasks in terms of the quality of the task product, the incurred resource cost of production, and user acceptability. The introduction of CAD technology in the office (Chapter 3) is intended to save costs and improve design quality. The system, therefore, could be considered effective if cost-benefits are realised and if it is accepted by users.

Product quality

In this research, *quality* refers to accuracy of the achieved drawing product, determined from the number of errors as defined below. (Quality will be used interchangeably with accuracy, as will product with task output.)

An *error* is defined as a deviation from an expected result or outcome. Errors relate to inaccurate drawing entities, assessed from a hardcopy output of the drawing. An *entity* is the smallest recognizable geometric figure which, if broken down any further, would be a set of points (Sharit & Cuomo, 1988). Examples of entities are lines, circles, arcs, ellipses and rectangles. The index for quality is stated as:

$$\textbf{Product quality} = \frac{1}{\Sigma\ (\mathrm{Pe})}$$

where Σ is the sum of

Pe is the number of errors in the drawing, as detailed below:-

- drawing elements that are crooked, disjointed entities, non-aligned entities, etc.;
- misplacement of objects within drawing limits relative to location in plan, and relative between one object and another;
- displacement of objects outside drawing limits, that is, objects positioned off-screen;
- wrong size of object relative to other objects in the drawing (eg. chair too large for desk, etc);
- inaccurate shape of objects (eg. ellipse looking like a circle, etc.).

Analysis of errors will be made by comparing the task product with the drawing plan using 'approximation judgement'. Therefore, the lower the index value, the higher the quality.

Production cost

Production costs, in this research, relate to user resources recruited to production (ie. for generating a proportion of the output), determined from measures of **time** and **efficiency**. (The term *efficiency* refers to effective use of resources.) Production cost (time) is measured as time per entity drawn; while production cost (efficiency) is measured as number of commands/data per entity drawn (ie. a measure of the efficient use of commands/data). Reference to these costs will be made as production time and production efficiency, respectively.

$$\textbf{Production time} = \frac{Pd}{Pa}$$

where Pd is set task time that is determined for task completion, measured in seconds.

Pa is the number of drawing entities achieved in the set time.

Therefore, the lower the index value, the lower the production time cost to the user.

$$\textbf{Production efficiency} = \frac{Pc}{Pa}$$

where Pc is the number of command/data entries in the <u>scored</u> time.

Pa is the number of drawing entities achieved in the <u>set</u> time.

Therefore, the lower the index value, the higher the production efficiency. (Note: It is not possible to obtain the number of entities in the *scored* time nor the number of commands in the *set* time, thus this value would result in commands being not whole commands.)

User acceptability

Performance is also assessed from the affective costs incurred. Here, the costs are determined from subjective measures of user acceptability. It is well-known that a system that is effective in supporting performance is likely to bring satisfaction to the user, tends to be rated favourably, and is preferred by the user (eg. Bailey, 1982; Eason, 1988; Booth, 1989). In other words, the system is accepted by the user. Users are known to have certain expectations of systems, for example, in using a speech recogniser. In instances where system performance tends to fall short of expectations, this creates a feeling of dissatisfaction, and consequently, the system gets rated poorly by the user.

The indices for determining acceptability are:

(1) *user satisfaction* - the affective state of being content with the performance attained, reflecting some of the emotion attached to using the system;

(2) *perceived performance* - the user's rating on the level of perceived achievement as

perceived, thus reflecting the degree of success in using the system; and

(3) *user preference* - the explicit choice for one system as opposed to another, reflecting desirability for a particular system.

This subjective assessment of user acceptability will be used to supplement objective assessments of system performance. However, it should be cautioned here that objective measures of performance may not necessarily correlate with subjective ratings of performance because of the complex interplay of factors such as motivation, learning, etc. Nevertheless, user acceptability measures are useful in determining the extent to which the system is well received by the user. These assessments will be obtained using interviews and structured questionnaires.

4.5.3 Behaviour-performance distinction

Distinguishing measures of behaviour from performance enables two types of statistical analysis to be carried out. Firstly, separate analyses can be made of the effects of different types of system on the variables that characterise behaviour and performance. This is achieved by using analysis of variance (ANOVA) that allows determination of cause-effect relationships. Such analyses provide comparative information regarding the strengths and limitations of each device or system type.

Second, an analysis can be made of the relationship between behaviour and performance measures using correlation analyses. This enables an assessment of the association between these variables. Thus, a positive correlation between a behavioural measure (eg. high frequency of eye transitions to text screen) and a performance measure (eg. high production time) would indicate that frequent gazing to the text screen may be associated with an increase in drawing time per entity. A negative correlation of similar measures would imply that a consequence of frequent gazing to the text screen is associated with increased productivity. Thus, correlation assessment would provide the basis for speculation regarding the interdependency of the individual variables, the extent to which the phenomena occur together, and the strength of the correlated variables.

Descriptive analyses of behaviour data (eg. mean analysis, percentage, etc.) provide the source for constructing a model of system behaviour. All statistical analyses are performed using the PC version of SPSS (see SPSS/PC+ V2.0 Base Manual, 1988).

4.6 SYSTEM BEHAVIOUR MODELLING

The framework selected here for developing a system model of design behaviour is the Blackboard model of design by Whitefield (1986b), which is derived from the HEARSAY model for speech understanding (Erman & Lesser, 1980). The model is developed post hoc to the research. Thus, its role is only analytical and will be used to explain the variability in performance between different CAD systems. The model is described in detail in Chapter 5.

4.7 HUMAN FACTORS GUIDELINES

Findings from this research will be used to develop guidelines for integrating speech and manual input in CAD systems. These guidelines will be accumulated over the experimental investigations and will be expressed in a standard format, following Smith and Mosier (1986). The form in which the guidelines are expressed is only formal in the sense of being as explicit and clearly defined as possible. Evidence from the literature will be used to provide further empirical support.

The guidelines might be used in two ways. First, by end users and CAD implementors in configuring multimodal CAD systems, using existing equipment and available technology. This would serve as a potential solution for users who are not willing or able to invest in new equipment. Second, by system designers in designing novel CAD systems that integrate speech and manual input. (The term *system designers* is used here in a broad sense to include system developers, engineers, human factors practitioners and design consultants.) In this context, the guidelines could serve as a design tool in system development. To ensure that the guidelines are usable, they will be validated by system designers in the workplace. Development and validation of the guidelines is described in Chapter 11. This, then, concludes the description of the methodology.

4.8 SUMMARY

This chapter describes a methodology for empirical evaluation of CAD systems, based on analyses of behaviour protocols and performance indices. The analysis is made possible by computerised techniques of scoring and data processing offered by VITAS. The purpose of the methodology is to afford an understanding of the constraints and limitations of CAD systems in supporting optimal draughting behaviour. Operational definitions of terms used in the thesis, particularly those related to optimal behaviour and performance, are also presented.

NEXT CHAPTER HIGHLIGHTS

One of the requirements of the above methodology is to develop a technique for modelling system behaviour. Chapter 5 presents a framework that would enable this requirement to be met. The framework is used to develop a blackboard model of system behaviour. Data for the model are derived from the empirical investigations to be described in subsequent chapters.

CHAPTER 5

A Blackboard Framework for Modelling System Behaviour

Overview

5.1 Introduction

5.1.1 Characterising user model

5.2 Modelling behaviour and performance

5.2.1 Normative-performative behaviour distinction

5.2.2 Model description - a summary

5.3 The framework for the model - the blackboard architecture

5.3.1 Blackboard concept

5.3.2 Knowledge notation as production rules

5.3.3 Blackboard control structure

5.4 Example blackboard system: Whitefield's blackboard model of design

5.4.1 Levels of design process representation

5.5 Blackboard architecture for modelling system behaviour

5.5.1 Assumptions

5.5.2 Relationship of system behaviour model to real user behaviour

5.5.3 A note about spatial (movement) dimension

5.6 Constructing a blackboard model of system behaviour

5.6.1 Data analysis

5.6.2 Model representation of system behaviour

5.7 Notional mapping of design model to system behaviour model

5.8 Summary

Next chapter highlights

CHAPTER 5

A Blackboard Framework for Modelling System Behaviour

OVERVIEW

A framework for constructing a model of system behaviour is presented in this chapter. The framework is based on the blackboard metaphor which has its origin in the HEARSAY speech understanding system. The model illustrates how behavioural knowledge is used in the performance of CAD tasks. The knowledge is not concerned with which input devices to use in carrying out the tasks, but more specifically, with how to operate a given device. Thus, the model is a description of actions and movements as defined in Chapter 4. The purpose of modelling system behaviour is to understand how well different input devices support task performance, specifically in reducing non-optimal behaviour. Such information could be used in optimising the design of CAD interfaces.

5.1 INTRODUCTION

The use of analytic techniques, such as user models, to model the interaction between users and computers is common in HCI research. The next section discusses user models.

5.1.1 Characterising user model

The idea of a user model is discussed at great length in the literature on human-computer interaction and intelligent knowledge-based systems (eg. Rich, 1983; Christie, 1985; Norman, 1986; Farooq & Dominick, 1988; Rivers, 1989). Murray (1988) reports ten different definitions. In view of the diversity in concep-tualising user models, Long (1987b), in an attempt to provide some coherence and clarity to the problem, described a framework for user models. For Long, a distinction between user models might be achieved by: (1) describing differently the two types of user models (*users'* models as mental or conceptual models held by users and *users* models as models of users by others); and (2) motivating them separately. For example, when users are modelled by 'paradigm agents', the models derived might be termed *agents' models of users*.

In general, models are representations of systems that specify the major components involved and the relationships among them. The representations are assertions about the properties of some entity which can be used in reasoning about that entity. The first step in model building, then, is to identify and represent those properties (eg. goals, behaviour). The second step is to incorporate capabilities for reasoning about them. It is generally agreed (eg. Booth, 1989; Young, 1985) that user modelling can:

- help in matching facilities that a system provides to the needs of the user;
- suggest metaphors to improve user learning (eg. desktop metaphor);
- guide design decisions and make design choices and assumptions explicit;
- provide a predictive evaluation of proposed designs;
- help to identify variations in the user population; and
- guide the design of experiments and help in the interpretation of results.

In other words, user modelling provides a mechanism for understanding much research in HCI.

Approaches to modelling vary according to the purpose of the model (eg. descriptive, prescriptive, predictive) and the content of the model (eg. processes, structure, strategy). Williges (1987) classifies approaches into two broad categories: conceptual and quantitative. The role of conceptual models is to represent cognitive knowledge in terms of *processes*, *structure* or *strategy*. The role of quantitative models is to represent performance knowledge numerically using empirical, simulation or statistical techniques. Therefore, models provide abstract, shorthand representations of users at some levels of conceptual or physical phenomena.

Most user models have been specifically crafted for each application. To achieve generality some characteristics are desirable in general user modelling. These relate to dimensions for measuring the generality of a user model. Kass and Finin (1988) suggest that user models may be general with respect to three dimensions: (1) the range of users; (2) the forms of interaction; and (3) the underlying system domain. User generality is a requirement of any user modelling facility, particularly if the system strives to adapt its behaviour to individual users. A user model has interaction generality if it can be used: (a) with a variety of interaction styles (eg. structured dialogues, mixed initiative dialogues); and (b) with various modes of communication (eg. menus, speech, graphics). A user model that is domain general can be applied in several contexts.

5.2 MODELLING BEHAVIOUR AND PERFORMANCE

A user communicating with a computer initiates an interaction between them; each has its own role to play and each has its own patterns of behaviour. The product of this interaction is behaviour related by performance to the achievement of a task goal (see Chapters 1 and 4). Models of system behaviour and system performance are generally formulated to serve one of two functions, either to predict behaviour and/or performance, or to compare the behaviour and/or performance of different system components.

The domain of predictive behaviour/performance modelling varies from models of specific component processes such as keystroke-level analysis (eg. Card, Moran & Newell, 1980), workload (eg. Wickens, 1983) to general models of the entire system, such as the GOMS model by Card et al. (1983). The domain of comparative behaviour/performance modelling is concerned with, for example, making a distinction between experts and novices (see Allwood, 1986), or between effectiveness of alternative input devices (see Chapter 2). Such comparisons enable identification of specific differences (eg. work style, quality of output) which could help in: (1) understanding behaviour/performance variability; and (2) suggesting ways of improving behaviour/performance through training, system design and the like.

Formal descriptions of system behaviour (eg. Moran's Command Language Grammar, 1981; Kieras and Polson's Formal Analysis, 1985) have recently gained popularity as a means of describing how the user's task maps on to the system's. While there are clear advantages to such formal specifications in terms of providing an explicit grammar for describing the user's tasks at the interface, formal modelling may not be appropriate for use in this thesis.

5.2.1 Normative-performative behaviour distinction

Often, a model of idealised optimal behaviour is devised with which performance of either the naive, novice or expert can be compared. The basis for such a model is ideal (or normative) expectancies, and hence is termed here a *normative* model. For example, a skilled typist, ideally, is expected to use all fingers in text-typing. However, observation of actual behaviour might reveal some discrepancy from the ideal. For example, the skilled typist is actually observed to use most or some fingers in typing. As Wickens (1984, p. 10) noted, "...it is important to realize that expert performance may indeed be far from perfect". The representation of actual behaviour constitutes what is termed here a *performative* model. This distinction between normative and performative behaviour is useful as it allows a comparison between expected and observed. Where there are discrepancies, attention may be focused on the nature of the differences, and solutions could be considered to bridge the gap.

Performative models, as explained above, reflect users' actual behaviour; the latter is, perhaps, guided by a set of reasonable rules (knowledge) that might look very different from those postulated by the researcher(s). Empirically, it has been shown that users tend to have various knowledge sources to rely on when using a system, and at times their knowledge is incomplete (Buckley & Long, 1987). Also, some skills (eg. word processing, graphics manipulation) require practice and learning. These skills rely on memory, specifically remembering what to do and how to act. In other words, performance variability is associated with both memory and motor control processes. In order to 'manage' performance, there is general agreement (eg. Kerr, 1983; Allport, 1980) that there is a need for constraints. These constraints (eg. goals, prior experience and learning, expectancy, order and completion biases) help to organise behaviour, promote efficiency and protect against errors.

5.2.2 Model description - a summary

The model to be described is developed within a research enterprise, in the domain of system performance engineering, for application in the CAD domain. Following Whitefield's (1986d) taxonomy, the modelling agent here is the *researcher*, and the modelled subject is the *system* (ie. user interacting with CAD software plus hardware to perform 2D draughting tasks). The model's content is *design behaviour knowledge*, derived from analyses of behaviour protocols of CAD users with varying levels of skill. In accordance with Long's (1987b) user model framework, the model might be aptly described as a *research ergonomist's model of system* to associate it with the ergonomics (engineering) paradigm. The model employs the blackboard framework, hence, it is a *blackboard model of system behaviour*.

To achieve generality, the model accumulates knowledge over a range of users and interaction modes. This generates two types of models within a space of user models:

- a model of a single, canonical user (or individual model), and
- a cumulative model of individual users (or group model).

This enables two types of knowledge recruitment comparisons: first, between individuals within a group (eg. same CAD systems or input devices but different CAD tasks or skills); and second, between groups summed across all users within the group (eg. different input devices but same CAD task or skills).

As mentioned in Chapter 4, the role of the model is descriptive and explanatory. The model will be used in the thesis in two ways. First, to relate the problem of non-optimal behaviour to the solutions under investigation by comparing between normative and performative behaviours. Any mismatch between expected and observed behaviours will be expressed in terms of the model. Second, to serve as a basis for developing the guidelines (see Chapters 4 and 11).

5.3 THE FRAMEWORK FOR THE MODEL - THE BLACKBOARD ARCHITECTURE

This section introduces the basic architecture of blackboard systems. The description will be in three parts. The first part presents the blackboard concept; the second part discusses knowledge notation; and the third examines control mechanisms in blackboard models.

5.3.1 The blackboard concept

The idea that behaviour is a hierarchical system of functions has parallels with the notion of a blackboard, a hierarchically structured, global database divided into parts and subparts. The blackboard concept is due to Newell (1969), and later reinterpreted by Simon (see Davis & King, 1977) who suggested it to the designers of the HEARSAY speech understanding system. HEARSAY-II was the first blackboard system to be constructed (Erman & Lesser, 1980).

The architecture is becoming increasingly popular as a method for the construction of systems which operate in domains requiring different kinds of knowledge to be applied in order to arrive at a solution to a problem (Craig, 1988; Ablett, 1988). CAD is one such domain which involves the recruitment of qualitatively different types of knowledge in performing the task (Whitefield & Warren, 1989). Thus, using the blackboard framework to model system behaviour seems applicable. Moreover, the architecture is a comparatively informal construct which lends itself to a variety of applications for solving problems in diverse tasks, such as psychological modelling (eg. Rumelhart, 1976), errand planning (eg. Hayes-Roth & Hayes-Roth, 1979), and robotics (eg. Velthuijsen, Lippolt & Vonk, 1987). Although these models differ in terms of structure and content, the commonality in the way the architecture is organised enables comparisons across models.

The blackboard architecture is based on the salient features of the HEARSAY-II system. Here, the major features of the architecture will be described briefly. There are three main components of blackboard systems:

(1) a globally accessible database called the *blackboard*. The blackboard is structured as a hierarchy of abstraction levels. The blackboard contains the results of applying problem-solving knowledge;

(2) a set of *knowledge sources* (KSs). Knowledge sources represent the problem-solving knowledge contained in a blackboard system. They respond to changes in the state of the blackboard database by altering its contents;

(3) a control component called the *scheduler*. The scheduler is charged with controlling the problem-solving behaviour of the system. It does this by monitoring the current state of the blackboard and selecting one or more KSs to apply to it.

Each of these components will be explained in brief. The blackboard records the current state of the problem solution. The items placed on the blackboard are called *entries*. Entries represent elements of the solution being developed. An entry may be linked to other entries on the same, on higher, or on lower abstraction levels to form solution islands. The links are organised as an AND graph. The purpose of the link structure is to assemble individual entries into collections which constitute potential solutions.

The blackboard is used by all KSs as their only means of communication. KSs may only communicate by modifying items already on the blackboard or by adding new items. They may incorporate complex procedures for bringing about changes to the current state. KSs are independent in that they do not invoke one another and have no knowledge of each other's behaviour or existence. But they are cooperative in that they contribute solution elements to a shared problem. By permitting the KSs to influence one another's problem solving behaviour only indirectly, the architecture is able to achieve simultaneous independence and cooperation among KSs.

KSs are structured as condition-action pairs. The condition-part monitors the blackboard for changes such as the addition of a new entry or the modification of one already present. Such a change is often referred to as a *blackboard event*. The action-part of a KS makes changes to the blackboard. It can add or modify entries. When the condition-part of a KS has been satisfied, the KS is usually referred to as having been triggered. A triggered KS (TKS) represents a unique triggering of a particular KS by a particular blackboard event. A TKS will not be executed unless it has been chosen by the scheduler. If it is selected, its KS's action executes in the context of its triggering information, and produces new blackboard events.

In general, more than one trigger will be applicable, that is, more than one KS could be activated given the current state. The scheduler is responsible for implementing one or more problem-solving strategies which guide the system's problem-solving activity. It acts upon TKSs and uses problem-specific parameters to determine which KS action to execute. The scheduler examines the blackboard state and those KSs which have been triggered, and makes its selection on the basis of the prescriptions of the strategy currently in force. The control task consists simply of taking the first element from the ordered list (schedule) maintained by the scheduler and executing it.

Having described the basic concept, the notation system and control structure in blackboard architecture will now be described.

5.3.2 Knowledge notation as production rules

A notation is a representation for describing some aspect of a system or user behaviour. The framework described above uses production rules as a formalism in knowledge representation. The use of productions has been the subject of much work since their introduction by Post (1943) as a general computational mechanism. They have been used to describe various cognitive processes, including problem-solving (eg. Newell & Simon, 1972), learning (eg. Anderson, 1983; Kieras & Polson, 1985) and complex human-computer interactions (Durett & Stimmel, 1982).

A production system, according to Barr and Feigenbaum (1983) consists of three parts: (1) a *rule base* composed of a set of production rules; (2) a special, buffer-like data structure called the *context*; and (3) an *interpreter* which controls the system's activity. Its utility and limitations in knowledge representation has been much discussed in the literature (eg. Davis & Lenat, 1982; Barr & Feigenbaum, 1983). One advantage of production systems is that the individual productions in the rule base can be added or modified independently. Another is the uniform structure imposed on the knowledge in the rule base. A further advantage is the ease with which the knowledge may be expressed. There are, however, disadvantages inherent in the production-system formalism. One of these is inefficiency of knowledge execution, resulting in high computational overhead in their use (Barr & Feigenbaum, 1983).

5.3.3 Blackboard control structure

Most blackboard models differ in terms of their control architecture. In the HEARSAY-II system, scheduling and activation of KSs is based on a complex focus-of-attention strategy. The priority of a KS waiting for execution is based on principles such as best-first, validity, significance, efficiency and goal satisfaction (see Hayes-Roth & Lesser, 1977). In Whitefield's (1986b) design model, control of KSs is based on meta-level knowledge, that is, knowledge about knowledge (Davis, 1980). This meta-knowledge is represented and applied in the same way as the object-level KSs (ie. knowledge of the task domain), but is not written to the blackboard.

Warren (1987) in applying Whitefield's framework to another engineering design domain (pipework), employed a more detailed control structure for scheduling object-level KSs. This incorporated a scheduler which relied on control heuristics in the form of meta-rules to order and schedule the activated KSs. Meta-rules embody strategies, that is, knowledge that indicates how to use other knowledge (Davis & Buchanan, 1977). The concept of strategies as a mechanism for deciding which knowledge to invoke next has been applied to different control structures, for example, in Hayes-Roth's (1985) blackboard for control, Lesser and Corkill's (1981) sophisticated scheduler, and Lenat et al.'s (1983) meta-level architecture.

The quality of a knowledge base depends not only on how well it solves problems, but also on how easily its design allows it to maintain and modify the knowledge base without extensive effort. This requires that control knowledge be represented abstractly, separate from the domain knowledge it operates upon (Clancey, 1983; Wilkins, Clancey and Buchanan, 1987).

As an illustration of the features described above, the following section describes, in brief, a blackboard model of design by Whitefield (1986b).

5.4 EXAMPLE BLACKBOARD SYSTEM: WHITEFIELD'S BLACKBOARD MODEL OF DESIGN

This description of Whitefield's (1986b) model of design is necessary as a basis: (1) for comparing between different blackboard models; and more specifically, (2) for constructing this model of system behaviour.

The framework in Whitefield's design blackboard model is a set of assumptions concerning process and structure that defines a space of possible models. The details of particular models depend upon empirical data derived from analyses of verbal protocols of designers with and without CAD. The blackboard is where the design solution is constructed. It is divided into levels, each of which contains a description of the solution, but at different levels of abstraction termed *unit, item* and *detail* levels.

The object-level KSs comprise those that are concerned with domain knowledge (in this case, instrument casings), called *domain KSs*, and those related to various aspects of drawing and system operation, termed *drawing KSs*. The KSs are divided into those that *generate* solution elements and those that *evaluate* existing solution elements. The order in which these object-level KSs act is controlled by meta-level knowledge.

Comparatively, it could be said that Whitefield's blackboard design model differs from the HEARSAY-II blackboard model, both in *quantity* (eg. number of levels) and in *form* (eg. content). The blackboard in HEARSAY-II is a generalised three-dimensional network, with information level (partitioned into six levels of the speech signal), time within an utterance, and competing alternative hypotheses. The design blackboard has two dimensions, namely, information level (partitioned into three levels of the object), and solution within a space. This difference in dimensions is one of quantity. The content of HEARSAY-II is about speech utterance, while in the design model the content is about design knowledge. This difference is one of form.

5.4.1 Levels of design process representation

Both blackboard design models (see Whitefield & Warren, 1989) are concerned with object-level KSs at a high level of task execution. In other words, these models specifically address knowledge recruitment at the *task* level, with less emphasis on knowledge recruited at the lower level of *input/output*. Using a similar framework to construct a model of system behaviour for the purposes of this thesis, it is essential to describe knowledge that is concerned primarily with I/O. These KSs relate to actions and movements and are called here *behavioural knowledge*.

So, the purpose of this system behaviour model is to describe the behaviour and to explain how different behaviour KSs are recruited in the interactive process. Given that this model is a modification of the design blackboard model, it assumes similar structure, with regard to the number of levels on the behaviour domain blackboard. This model's assumptions, however, are based on Hayes-Roth's (1985) control blackboard architecture. As such, this model to be developed has some differences from Whitefield's in terms of its content and overall architecture.

5.5 BLACKBOARD ARCHITECTURE FOR MODELLING SYSTEM BEHAVIOUR

In this section, a framework for constructing a model of system behaviour is described.

5.5.1 Assumptions

Adapting the assumptions of Hayes-Roth's (1985) framework to this domain, they read as:

Assumption 1

> *All solution elements generated during problem solving will be recorded in a structured, global database called the Blackboard. The blackboard architecture defines an explicit domain blackboard and control structure.*

The blackboard structure organises behaviour (action and movement) solution elements along two axes, solution intervals and levels of abstraction. Solution intervals represent **time** on the *temporal* dimension, and **movement space** on the *spatial* dimension. Levels of abstraction represent interaction entities, namely, SubTask, Action, and Movement on the *interaction* dimension (or STAM abstraction levels). The Subtask level represents subgoals; the Action level represents molar behavioural operations (or gross actions); while the Movement level represents molecular operations (or fine movements). Derivation of these blackboard levels is predetermined, based on typical hierarchic breakdown of behavioural functioning units, as generally described in the literature on motor behaviour and human performance (eg. Schmidt, 1983; Kelso, 1982; Singer, 1980). Thus, this assumes that the minimum number of levels in the model will be three.

The behaviour domain blackboard records solution elements for the current CAD problem (eg. entering a command, creating an object entity, etc.), but not the executed actions. Its solution intervals and STAM abstraction levels are domain-specific, while its behaviour description is determined, in part, by the systems technology. The control structure orders the sequence of triggered KSs and executes their actions based on heuristics in the form of meta-rules. The rules embody criteria which would contribute to optimal behaviour, leading to an enhancement in task performance.

Assumption 2

> *Solution elements are generated and recorded on the blackboard by independent processes called knowledge sources. The blackboard architecture defines explicit domain and control KSs.*

Knowledge elements are represented as condition-action units in the form of production rules, expressed in multiple IF-THEN format. The use of a multiple format enables better organisation of KSs, which could lead to a more efficient firing of rules, and to fewer production counts. The condition (IF) describes situations in which the KS can contribute to the performing process. Ordinarily, it requires a particular configuration of solution elements on the blackboard. The action (THEN) specifies the KS's behaviour. Generally, it entails the creation or modification of solution elements on the blackboard. Only KSs whose conditions are specified can perform their action. A description of a generic production notation is shown in Figure 5.1.

IF

[TASK] SUBGOAL = description of subgoal
[COMPUTER] INPUT/OUTPUT DEVICE = description of input/output device
[USER] SKILL = description of user capability

THEN

[VISUAL] Look-Target (gaze at computer, task tools, non-task targets)
[MANUAL] Key-in-Information (depress key/transducer)
[VERBAL] Verbalise-in-Information (speak into speech recogniser)
[COGNITIVE] Remember-Information

Figure 5.1 A generic production showing the possible conditions and actions

The list of conditions is prefaced with two or three associative IFs which specify that the production will fire if all of the conditions in the associatives are true. Each of the conditions is a specification of system components relating to Task (subgoal), Computer (input/output device) and User (user skill). The action is a sequence of operations that modifies the contents of the blackboard, and generates user actions that are implemented either as *gross action* outputs (eg. eye gaze, hand transition, speech utterance, etc.), or *fine movements* (eg. finger press, eye saccade, etc.). The actions and movements implemented constitute real world behaviour, and they cause the change in the state of the display or other information to which the production system attends.

User skill refers to the competence (capability) that users have in performing sets of tasks ascribed to them. A broad view of skill would include experience, ability to learn, coordinative ability, work style, values, etc. This skill will determine how and why the user performs in a certain way.

Each KS has attributes which characterise it. The attributes are:

- **Name**, an identifying label
- **Description**, a statement of its characteristic behaviour and functionality
- **Condition**, a situation of interest for trigger
- **Action**, a series of blackboard changes

A complete listing of behaviour KSs is given in Appendix 2.

Identifying Behaviour KSs

Behaviour KSs operate on the domain blackboard at each level of the blackboard. These KSs are divided into domain and drawing KSs, similar to those of Whitefield's (1986c); however, they are called here *Task Specific* and *Tool Management* KSs, respectively. Within each KS type, there are two types of KS that have the functions of: (1) generating solution elements between and within blackboard levels, hence called **Generative** KSs (these have similar functions to Whitefield's); and (2) instructing or directing which solution element to use within the blackboard levels, called **Instructive** KSs.

Task Specific KSs

Unlike Whitefield's domain KSs, these Task Specific KSs operate between the lower level of the 'Communications' blackboard (see its notional representation in Section 5.7) and the SubTask level of the I/O blackboard. They consist mainly of *subgoals*, underlying the reason for doing the subtasks, including the type of information input, the task object and instructions. Examples of Task Specific KSs are creating an object "desk" quickly, modifying an object entity "line" accurately, entering a command "rectang:", specifying the parameter of an object "25mm radius", etc.

Generative Task Specific KSs are:

- *types of information* (eg. commands, coordinates, numeric, text);
- *types of object* (eg. desk, chair, sink, bath, etc.);
- *types of task instruction* (eg. accuracy, speed); and
- *types of menu* (eg. tablet menu 1, tablet menu 2, etc.).

Instructive Task Specific KSs are:

- *types of commands* (eg. DRAW [line or rectang], EDIT [erase last or erase window], DISPLAY [zoom all or zoom window], etc.); and
- *types of object parameters* (eg. length of chord/line [5 mm or 5.5 mm], radius of circle [120 mm or 110 mm], direction of angle [45 degrees or 135 degrees], coordinate position [45,60 or 45,70], etc.).

Tool Management KSs

These KSs operate within and between all levels of the blackboard. They comprise the *system component tools*, underlying the I/O means for communication. These encompass user, computer, task and non-task tools (termed here *elsewhere*).

Generative Tool Management KSs are:

- *user tools* (eg. eyes, hands, voice, etc.);
- *screens/output devices* (eg. graphics screen, text screen, plotter, printer, etc.);
- *input devices* (eg. graphics tablet, keyboard, speech recogniser, etc.); transducers (eg. stylus, puck, etc.);
- *task objects* (eg. plan, manual, calculator, command list, etc.); and
- *elsewhere* (eg. mug, colleague, etc.).

Instructive Tool Management KSs are:

- *hand type* (eg. right hand, left hand or both hands);
- *finger type* (eg. all fingers, some fingers or middle finger, etc.);
- *menu item type* (eg. tablet menu item or screen menu item);
- *prompt type* (eg. graphics prompts or text prompts);
- *vocabulary list type* (eg. online speech list or offline speech list);
- *transducer key type* (eg. puck key or keyboard function key, etc.);
- *calculator type* (eg. portable calculator or on-screen calculator); and
- *writing object type* (eg. pencil or pen, etc.).

Behaviour KSs that do not write to the blackboard, but are responsible for the precise and fine tuning of blackboard entries, are those that relate to movement control. These KSs regulate and control the implementation of output by applying variable degree of force, pressure, timing, rhythm, volume, etc. To distinguish these from all classes of Task Specific and Tool Management KSs, such KSs will be called here *Movement Control* KSs.

Identifying Control KSs

Control KSs contain strategic knowledge in the form of meta-rules. The meta-rule has two parts: the first part is the description of conditions which can be global, local or current. The second part is the description of actions. The rules are judgmental, that is, they specify inexact interpretation, with the strength of the interpretation given on a scale from 0 to 1. The rules are invoked by an interpreter that is data-driven. Examples of meta-rules are:

(a) IF a KS is known to be more efficient than another KS
THEN select that KS

(b) IF a KS is operating on a part of the blackboard known to contain important information for achieving subgoal
THEN give it priority

(c) IF a KS is known to be frequently in use
THEN choose that KS

(d) IF an *eye-graphics screen* and *eye-text screen* KSs are competing,
AND the latter contains important information for interpreting input

THEN give the latter KS higher priority than the former

(e) IF an *eye-graphics tablet* KS is competing with a *hand-graphics tablet* KS
THEN give both KSs equal priority

(f) IF on the current subtask the KS associated with speech recogniser has been executed *n* times due to misrecognition of input,
AND IF on the current subtask there is a KS associated with graphics tablet as backup facility,
THEN the priority of the KS for graphics tablet should be increased

The difference between the above control KSs is that: the rules in (a) to (c) apply to all KSs, and are thus *global* rules, while the rules in (d) and (e) apply only to specific KSs, and are thus *local* rules. The rule in (f), however, is a local rule but applies specifically to a current problem situation, and is thus termed a *current* rule.

Assumption 3

On each problem solving cycle, a scheduling mechanism chooses a single KSAR to execute its action. The blackboard architecture defines a simple scheduling mechanism to manage both domain and control KSARs.

Using terminology from the domain of motor behaviour (Schmidt, 1982), the scheduler for managing behaviour and control KSs is called a *knowledge executor*, and the record of activated KSs is called an *action plan*, which has similar functions as the 'agenda' in Warren's (1987) scheduler. Like his scheduler, the knowledge executor has three parts: (1) an interpreter; (2) an action plan; and (3) a rule database.

Figure 5.2. illustrates the control architecture, mapping the knowledge executor to the domain blackboard rules. The interpreter performs the basic function of reading and writing control KSs from and to the action plan. The interpreter determines the order for TKS execution based on a system of weighting (ie. a rating on a scale of 0 to 1). The action plan contains control information, such as history of KS use; importance of KS to optimal behaviour; current subtask/problem indicator; priority of subgoal; and history of knowledge executor operational routines. The rule database contains control heuristics in the form of meta-rules at three different levels: global, local and current. Examples of each rule type are given above.

The control KSs iterate a three-step procedural cycle, on the action plan, as follows: (1) update-to-do-set; (2) choose-TKS; and (3) execute-chosen-TKS. The first procedure takes the events that occurred during the last completed cycle and asks the interpreter for those KSs that have a matching trigger. The returned KSs are given to the interpreter for creation of a knowledge source activation record (KSAR). A KSAR is an item on the action plan. A KSAR that has its conditions met will be triggered. The interpreter computes the priority of a

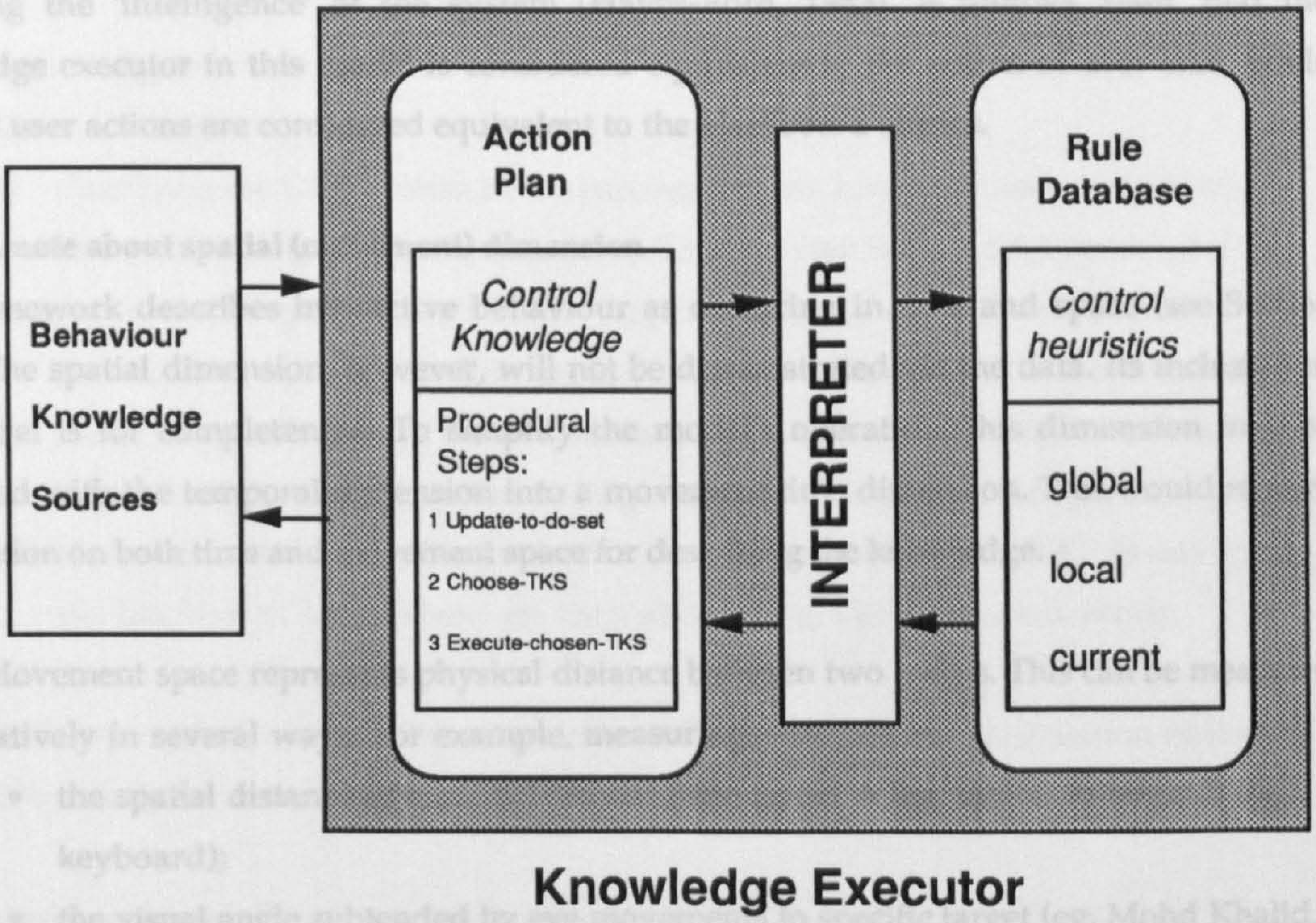


Figure 5.2 Control Structure Architecture: Knowledge Executor

triggered KSAR by applying domain-dependent rules. On completing this procedure, the second procedure decides which KSAR to be executed. The selected KSAR is placed at the top of the list of KSARs, and the third procedure completes the cycle by executing the action of the chosen KSAR.

5.5.2 Relationship of system behaviour model to real user behaviour

The model represents behaviour knowledge which is implemented as gross actions and fine movements. These actions and movements constitute real user behaviour. The scheduler (knowledge executor) which controls the operation of the behaviour KSs has been regarded as reflecting the 'intelligence' of the system (Hayes-Roth, 1983). It follows, then, that the knowledge executor in this model is considered equivalent to the notion of user skill, while planned user actions are considered equivalent to the blackboard entries.

5.5.3 A note about spatial (movement) dimension

The framework describes interactive behaviour as occurring in time and space (see Section 5.5.1). The spatial dimension, however, will not be demonstrated via the data. Its inclusion in the model is for completeness. To simplify the model's operation, this dimension may be combined with the temporal dimension into a movement time dimension. This would require information on both time and movement space for describing the knowledge.

Movement space represents physical distance between two points. This can be measured quantitatively in several ways. For example, measuring:

- the spatial distance of hand movement from target A (eg. tablet) to target B (eg. keyboard);
- the visual angle subtended by eye movements to specific target (eg. Mohd Khalid, 1981);
- the physical distance on screen between two objects or two coordinate points; or
- the waveform of speech utterance using spectral analysis technique (eg. Bailey, 1985), etc.

These measurements would provide fruitful information about various anatomical orientations (eg. head turning, body displacement) of the user during system use.

Spatial (movement) data are clearly a useful source of information for supporting system design, particularly in configuring device layout. For example, design of CAD interfaces might consider: (1) incorporating keyboard features as part of a standard graphics tablet, so as to confine hand transitions within a small range of the tablet; or (2) designing dual-screen systems that confine eye transitions to just upwards and downwards, and therefore involve small head movements, rather than the sideways eye movements, with larger head orientations, as incurred by current layout of display devices. Systems should be designed to minimise the effects of physical strain and stress on the users (Pheasant, 1988). This spatial data would thus be useful to system designers.

Having described the framework, the next section presents the construction of the model.

5.6 CONSTRUCTING A BLACKBOARD MODEL OF SYSTEM BEHAVIOUR

Using the framework described in Section 5.5, this section describes how the model will be developed.

5.6.1 Data analysis

The analysis involves:

(1) classifying the CAD system into components (user, computer, task and elsewhere). Then, classifying each component into KS types (see Section 5.5.1 for details) to generate a complete set of KSs (see Appendix 2). This classification produces a generic class of behaviour KSs applicable to human-computer CAD systems; and

(2) classifying each KS type into those that are Task Specific and those that concern Tool Management; within these categories identifying those that generate solution elements, and those that have the functions of instructing which KS to use within the blackboard level. These are then allocated to the blackboard levels.

This two-step process is meant to be exhaustive so that knowledge representation of the CAD task will be rendered complete at the relevant levels.

5.6.2 Model representation of system behaviour

Figure 5.3 is a representation of the system behaviour blackboard. The horizontal lines in Figure 5.3 represent the STAM abstraction levels of the blackboard. The circles represent the level of input to a KS and the arrows its level of output. The KSs operating at different levels of the blackboard are summarised in Table 5.1 (see Appendix 2 for a complete list). An overview of the model's operation is described below.

With reference to Figure 5.3 and Table 5.1, information elements from the notional Communications blackboard are placed on the SubTask level by Generative Task Specific KSs (Type A). These are then identified by Instructive Task Specific KSs (Type D) which direct the use of solution elements on the SubTask level. These elements are then read off from the SubTask level by Generative Tool Management KSs (Type B) which generate solution elements on the Action level of the blackboard. Solution elements already on the Action level are identified by Instructive Tool Management KSs (Type E) which instruct which solution

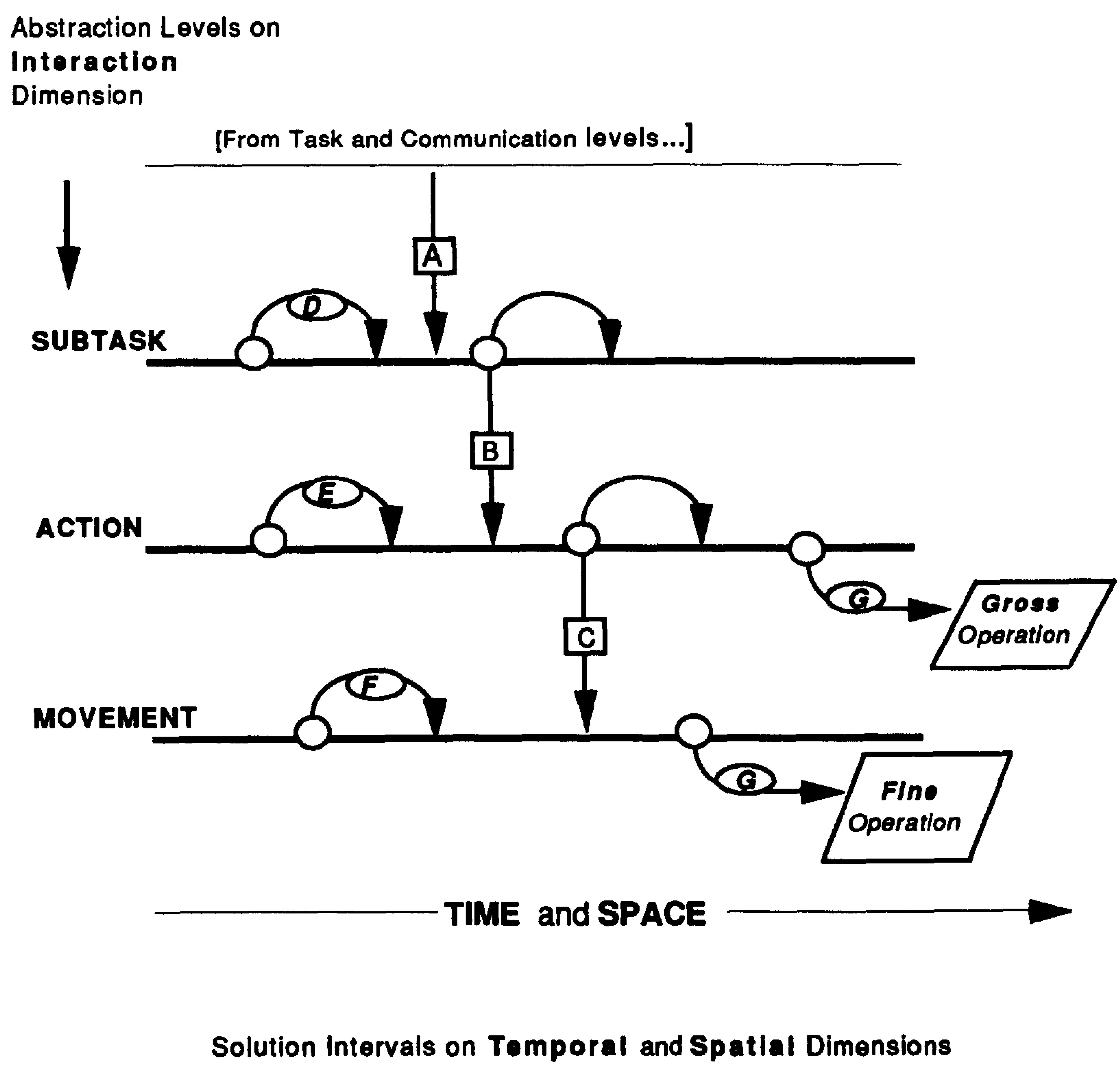


Figure 5.3. System Behaviour Blackboard Representation at Input/Output Level

TABLE 5.1

A Summary of Behaviour Knowledge Sources

Behaviour KS Type	*Knowledge sources*
A - Generative Task Specific :	Information type; Object types; Instruction types; Menu overlay types
B - Generative Tool Management :	**Sense receptors**; **Screens**; **Input devices**; **Transducers**; Task objects; Elsewhere
C - Generative Tool Management :	**Sense receptors**; **Screens**; **Input devices**; **Transducer** keys; Keyboard Keys
D - Instructive Task Specific :	Command types; Object parameters
E - Instructive Tool Management :	Right-hand; Left-hand; Both-hands; Graphics screen prompts; Text screen prompts; Tablet menu; Screen menu; Online vocabulary; offline vocabulary; Portable calculator; Computer calculator
F - Instructive Tool Management :	All-fingers; Some-fingers; Alphanumeric key; Function key; Puck key; Cursor key; Tablet menu item; Screen menu item
G - Movement Control :	Force; Stress; Timing; Pressure

element to use within the blackboard. These are placed on the blackboard and implemented as gross operations (eg. eye gaze; hand transition, etc.). Regulation of these outputs are made by Movement Control KSs (Type G) which read from the blackboard entries on the Action level, but do not write to the blackboard (eg. timing eye-hand coordination). The gross actions implemented constitute real world behaviour.

However, elements on the Action level can generate further solution elements on the Movement level of the blackboard (Type C). These are then identified by Instructive Tool Management KSs (Type F) which direct the production of fine movements (finger-press, etc.). Like gross actions, fine movements implementations are controlled by Movement Control KSs (Type G), and the outputs represent real behaviour at a molecular level.

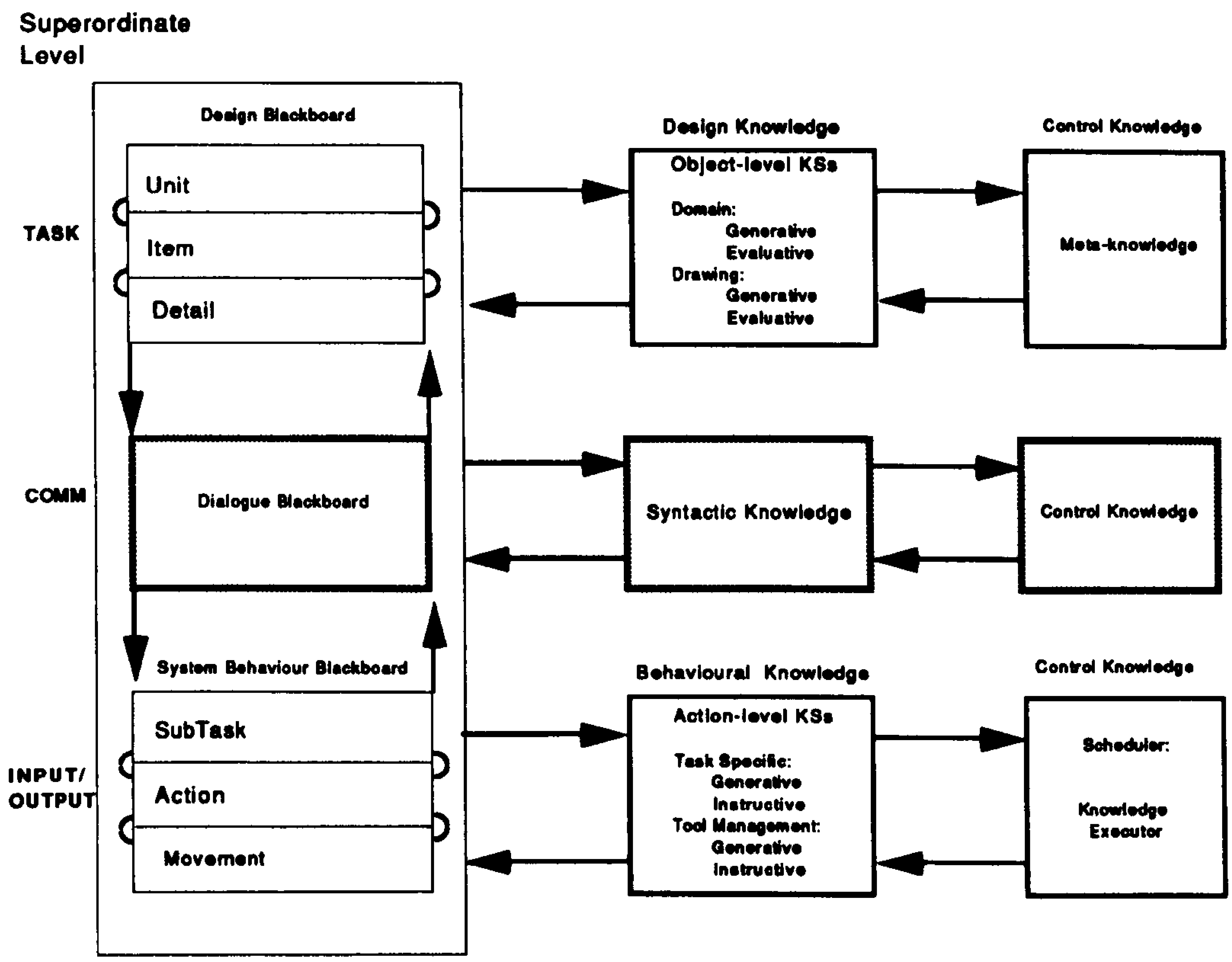
5.7 NOTIONAL MAPPING OF DESIGN MODEL TO SYSTEM BEHAVIOUR MODEL

Figure 5.4 shows a configuration of three blackboards at a superordinate level of the design process. At a high level is the task, represented in terms of a design blackboard (after Whitefield, 1986b). Besides task knowledge, users need to have knowledge about the functionality of the system with respect to the underlying task space. The functionality itself needs to be described and evoked by a suitable dialogue (eg. command language) which in turn is described in terms of its syntactic and semantic features. Such knowledge 'mediates' between task and I/O. Thus, it is represented here at an intermediate level as communication, in the form of a notional dialogue blackboard. At a lower level of the design process is I/O, represented by the system behaviour blackboard described here. This notional mapping is aimed to contextualise the behaviour model within an overall design process blackboard. This division of the superordinate design blackboard into three sub-blackboards at the task, communications and I/O levels is in accordance with Long's (1987d) framework.

5.8 SUMMARY

To summarise, the framework consists of a central blackboard, with dimensions of interaction, temporal and spatial levels, which is read from and written to by a variety of behaviour KSs. These KSs can be thought of as containing knowledge in the form of production rules. The blackboard contains hypothesised solution elements, with related elements linked together. The KSs scan the blackboard looking for the conditions that will evoke their actions of creating or modifying blackboard elements. The order in which the triggered KSs are allowed to act is controlled by control KSs in the form of meta-rules.

Within this architecture, several models of system behaviour could be developed. Using similar analysis techniques these models would enable within- and between-group comparisons of behaviour. The contents of the four models in this thesis are derived from analyses of



Notional mapping of 3 different blackboards at Task, Communication and Input/Output levels

Figure 5.4 Contextualisation of System Behaviour Blackboard within a Superordinate Design Process Blackboard behaviour protocols obtained in empirical studies from different CAD user groups. The analysis involves classifying each behaviour element into different KS types. These are then allocated to the blackboard levels. The KSs are general in the sense that they encompass generic objects of human-computer systems. In this respect, the knowledge might be generalisable to other domains of application (eg. word processing) which involve the use of similar system tools.

NEXT CHAPTER HIGHLIGHTS

Chapter 6 describes an observational study of CAD experts at work. The data will be used to construct models of system behaviour based on expert performance. Comparisons between individual models will be made in order to understand the recruitment of knowledge during CAD performance.

CHAPTER 6

Study 1: Observing CAD Experts at Work - Documenting the Problems of Manual Input

Overview

6.1 Introduction

6.2 Method

6.2.1 Description of systems studied

6.2.2 Procedure

6.3 Results

6.3.1 Characterising CAD expert behaviour: description of the group data

6.3.2 Characterising CAD task knowledge: comparisons of individual data

6.3.3 Analysis of questionnaire protocol

6.4 Discussion

6.4.1 Comparisons of individual models of system behaviour

6.4.2 Problems documented

6.4.3 Possible solution

6.4.4 Conclusion

6.5 Summary

Next chapter highlights

CHAPTER 6

Study 1: Observing CAD Experts at Work - Documenting the Problems of Manual Input

OVERVIEW

In Chapter 1, it was stated that the unitary use of manual input devices results in non-optimal behaviour. This chapter describes a field study aimed at identifying the nature of non-optimal behaviour in using the graphics tablet and keyboard. This documentation of the problem is based on a characterisation of visual and manual behaviour of CAD experts at work. The behaviour protocol is gathered through indirect observation and video-recording of the CAD activity as it occurs in the workplace. The protocol data are used to construct a model of system behaviour using the framework described in the previous chapter. An analysis of the problem will be made using: (1) the group and individual protocol data; and (2) data from the questionnaire.

6.1 INTRODUCTION

In Chapters 1 to 3, it was concluded that there is little empirical research which examines the behavioural aspects of input devices, particularly in the CAD task domain. Therefore, this study is essentially exploratory, to document the problems of manual input use in terms of:

- what proportion of the time on task is allocated to: (1) on-screen gazing; (2) off-screen gazing; and (3) operating input devices;
- what are the frequencies and duration of each behaviour type;
- when and why are the different system tools used; and
- how skilled are the designers in manipulating the input devices.

This information is crucial for determining whether CAD expert behaviour is optimal as defined in Chapter 4. This will be done by comparing normative (expected) behaviour with performative (observed) behaviour. Any differences will be explained in relation to the types of knowledge recruited during task performance. The identification of alternative solutions, such as the use of speech-based input devices, will depend on the nature and extent of the problem as documented here.

The technique for gathering the behaviour protocols is indirect observation. (Here, *direct observation* is understood as observation of behaviour in natural or laboratory settings performed with the aid of notes, checklists and without the use of audiovisual recording.

Indirect observation is understood as observation of such behaviour via audiovisual recording or video monitoring.) For most applications, indirect observation permits much more flexibility and accuracy in the observational methods - provided the behaviour of interest can be adequately captured on videotape. For any analysis which is aimed at exploring the structural aspects (ie. patterns) of interactive behaviour, an audiovisual recording would be useful (Clarke & Ellgring, 1983). This is the reason for the use of indirect observation methods in this exploratory study. The next section describes the method of data collection in more detail.

6.2 METHOD

The study comprised:

(1) a sampling of design activity in several interactive CAD systems. This was recorded on videotape, thus allowing detailed analyses of direction of eye gaze to specific targets of interest, and of movements of hands while operating the devices. These data are necessary for establishing that the unitary use of manual input produces non-optimal behaviour; and

(2) a subjective assessment of opinions concerning usability of CAD hardware and software, obtained by means of a questionnaire (see Appendix 3). A summary of the questionnaire data is given in Appendix 4. The data are required to supplement the behaviour protocol findings.

6.2.1 Description of systems studied

This section describes the CAD systems studied, in terms of the system components, that is, Computer (hardware and CAD software), User (designers), Task (design and digitising) and Environment (location).

Location

The study was conducted in the design departments of three prominent British engineering and petroleum-based companies, all located in London, England. Two of the companies have parent conglomerates overseas. The organisations are Fluor (Great Britain) Limited, British Petroleum International Limited and Bechtel Limited. These will be referred to as Organisation A, B and C, respectively. As private limited companies in a free enterprise economy, competition is an important element for economic survival. To remain competitive in their respective areas of specialization, utilising modern information technology is seen as crucial. To help increase productivity, CAD was introduced into the design offices to support the traditional drawing board activities.

Designers

Twelve male and three female CAD designers, aged between 18 and 50 years (mean=32.3

years), participated in the study. Three worked as cartographers, three as technician trainees, while the rest worked as design supervisors and senior designers. Twelve (80%) of the participants had used the conventional drawing board method before transferring to CAD systems, and nine (60%) had used their current CAD system for at least 3 years. As designers, 47% spent 9-10 hours continuously per day at the terminal while others worked between 7-8 hours per day, and generally 4-5 days a week. The long hours invested are perhaps explained, in part, by the competitiveness of the business enterprise.

CAD hardware and software

Fourteen workstations were observed (2 designers shared the same system). The designers in Organisation A used the GE CALMA Dimension III system which was operated from a mini computer. The designers in Organisations B and C employed the INTERGRAPH Interact 32C systems which were driven by VAX super-mini and mainframe computers. Both types of system used two-screen configurations. But the CALMA system had clearly separate alphanumeric and graphics terminals; the INTERGRAPH system had combined graphics and alphanumeric on each terminal.

The input devices used were the graphics tablet and keyboard. The keyboard had the QWERTY layout of alphanumeric characters with additional function keys designated to perform some CAD functions. The tablets were of two types: one a free-standing, portable pad with the stylus as the transducer; the other a desktop-tablet, integrated into the work surface with a cursor-puck as the transducer. The first type is used in the CALMA system, while the second type in the INTERGRAPH system. The pucks had 12 buttons, four of which were programmed to function as frequently-used CAD commands, namely, (1) enter command, (2) enter data point, (3) reset command, and (4) select tentative point.

The tablets were used with standard and tailor-made menu overlays. As many as three overlays are used at any one time, customised to fulfill specific design needs. The menus are organised according to frequency of use and importance of design information (eg. commands, prototype symbols, etc.), and are often colour-coded to assist in visual search. Screen menus were used only occasionally in view of space constraints on screen, particularly with large-scale drawings.

Both 2D draughting, 3D wireframe and solid modelling packages from different software developers were sampled. The softwares were supplied by the respective system vendors. But Organisation C, in particular, uses its own 3D modelling software (see CAE Bulletin, December 1986). The programs were applied to the design of objects from simple activity schedules and process flow charts to complex structures such as steelwork and piping. In cartography, for example, the programs are used to digitise landscape features from topographical or geological maps.

CAD task

The tasks observed were part of the designers' normal work; therefore, they were not experimental tasks prescribed by the researcher. These tasks were generally design tasks; only one subject performed a digitising task. The distinction between design and digitising was made in Chapter 3. To help identify which task to use in subsequent experiments, data from the digitising task will be compared with representative design tasks. This comparison will illustrate the differences in knowledge recruitment between tasks. However, for constructing a model of system behaviour, data from the digitising task will be excluded in order to preserve the homogeneity of the data set.

On the whole, the systems studied here can be said to be representative of existing CAD systems, as described in Chapter 3.

6.2.2 Procedure

All observations were recorded on video using two cameras. Two issues require consideration in recording interactive behaviour. First, the ability to obtain reliable and valid measurements. Second, the relative obtrusiveness of the observation technique. These issues govern the positioning of video cameras that would capture behaviour in full, while not being too obtrusive to the person being observed.

The first issue is resolved by placing one camera (positioned between the screens) directly facing the designer, so as to capture precise eye movements. These movements include: (1) those that involve major head orientations; and (2) those that do not involve any shifts in head orientation, that is, eye saccades. A second camera was located behind the designer, at approximately 45 degrees from subject's seating position, in order to capture hand movements and other behaviours. This recorded image produced a side profile of the user, thus providing an overall perspective of visual and manual behaviour. Using a vision mixer, separate images from the two cameras were mixed to produce a split-screen image on video.

The second issue concerning obtrusiveness is resolved, in part, through the task and the work target (task output) set for the designers by the supervisor. The CAD task, as described in Chapter 3, can be attention-demanding and is often claimed to be an 'eyes-hands busy' task (see Chapter 1). These task characteristics help to minimise subjects' awareness of the recording equipment in the work vicinity. Furthermore, the desire to meet the deadlines for specific task output encouraged the designers to carry on as they normally would.

Each designer was observed for 30 minutes. At the end of each recording session, a short questionnaire was administered to the subject. The role of this questionnaire is to provide information on previous design and CAD experience, qualifications and a subjective assessment of system use.

The behaviour protocols were analysed using VITAS (see Chapter 4). A 15-minute segment of each subject's recorded protocol was scored continuously. This segment of the videotape was selected after 5 minutes of recorded activity had elapsed. This selection should eliminate any potential technical irregularities that may have occured in the initial stage of the recording activity which could affect the quality of the recorded image (eg. adjustment of camera). An important criterion for scoring is to determine precisely when a particular action begins and ends in order to achieve accuracy and to minimise ambiguity concerning a behaviour action category. To this end, the scoring criterion was tested for each individual tape before the actual scoring process was carried out. Using various VITAS data processing programs (see Appendix 1), the frequency and duration of the actions scored were calculated, expressed as means and percentages.

The types of behaviour that were scored from the resulting tape are summarised in Table 6.1. However, not all the behaviour types are relevant for the problem analysis. Therefore, the required metrics derived from scoring these behaviours include:

(1) frequency and duration of eye gaze to graphics screen, text screen, graphics tablet (puck/stylus), keyboard, plan and calculator;
(2) frequency of eye transitions between pairs of these I/O devices;
(3) frequency and duration of hand manipulation of graphics tablet and keyboard; and
(4) frequency of hand transitions between input devices.

To recapitulate on the terms used here, ***eye gaze*** refers to eye fixation to a target; ***eye/hand transition*** refers to the movement of eye/hand between pairs of targets; and ***hand manipulation*** refers to the amount of hand use.

For this study, *frequency* refers to the total number of occurrences of each behaviour type, relative to other behaviours (ie. relative frequency); ***duration*** is the total time spent per behaviour type (ie. relative duration). Both frequency and duration will be expressed as relative percentages. The term ***worktime*** refers to the time on task within the observation period.

6.3 RESULTS

The results will be presented in three parts. The first part will focus on the group results (n=14). The second part will use the individual subject data to emphasise the differences

TABLE 6.1
VITAS Menu for Scoring Behaviour Protocol of CAD Experts

A. Visual Behaviour

Direction of Looking
Text Screen (Screen 1)
Graphics Screen (Screen 2)
Tablet Menu
Puck
Keyboard
Plan
Calculator
Command/Prompt line
Crosshair/Cursor position
Colleague
Elsewhere

B. Manual Behaviour

Which hand
Right
Left
Both

Hand movement
Gross
Fine

Destination of Movement
Tablet
Keyboard
Elsewhere
Plan
Text (Screen 1)
Graphics (Screen 2)
Puck
Stylus
Pen/Pencil
Calculator
Manual

What Object
Puck
Stylus
Keyboard
Plan
Pen/Pencil
Calculator
Finger

Local Activity
Function key
Alphanumeric key
Function/Alpha key
Command key
Object abandoned
Puck reading
Tablet surface

What Movement
Selecting
Pressing
Pointing
Pick up
Put down
Holding
Writing
Calculating
No movement
Dragging

Destination of Pointing
Tablet menu
Plan
Graphics (Screen2)
Text (Screen1)
Miscellaneous
Elsewhere

between CAD performance as a function of I/O devices and task type. The last part will examine users' responses to the questionnaire, in particular the use of system information and their ratings on device manipulation skill and system use.

6.3.1 Characterising CAD expert behaviour: description of the group data

ANOVA analyses of the effect of system type (separate graphics/text system and combined graphics/text system) on eye gaze to screens revealed no significant differences between the two systems (see Appendix 5). This suggests that both groups of designers generate similar patterns of eye gaze, irrespective of system type. From the data, it was clear that one screen was used more actively than the other during performance. This active screen is one which displayed the detail drawing, and will be termed here graphics screen (or screen 1). The second screen will be treated as equivalent to the text screen (or screen 2). This, then, enables the separate group data to be combined as one.

Following the metrics described in Section 6.2.2, the results will be discussed as follows:

Eye gaze to screens, input devices and task tools

These results are aimed at identifying the distribution of attention between individual components of the system. Figure 6.1 shows the percentage frequency and duration of eye gaze to specific targets of interest. The nodes (circles) represent the targets and the values given denote the frequency/duration of eye gaze to the target. The lines joining the nodes indicate bi-directional transitions between pairs of targets.

From Figure 6.1, users spent, on average, 67.7% of their observed worktime at the terminal looking at one or the other screen. In other words, the screens (graphics plus text) are gazed for a longer period of the time, relative to other targets. The frequency of the screens being looked at during task performance is 56.7%. In terms of individual terminals, the graphics screen is gazed at for longer period (49.6%) than the text screen (18.1%). Gazing at the graphics screen also occurs more frequently (37%) than the text screen (19.7%). This suggests that attention is divided between the two screens; of the two, the graphics screen receives much more attention.

The frequency of eye gaze to input devices (tablet and keyboard) is 29.8%. The duration spent looking at these devices, however, is only 16.3%. This means that the designers gazed at the tablet and keyboard for shorter periods of the time. Of the two input devices, the tablet is gazed at more frequently (21.3%) than the keyboard (8.5%). The time spent looking at the tablet is 11.2% compared with 5.1% for the keyboard. Once again, the results suggest that attention is divided between the input devices; of the two in use, the tablet receives greater attention than the keyboard.

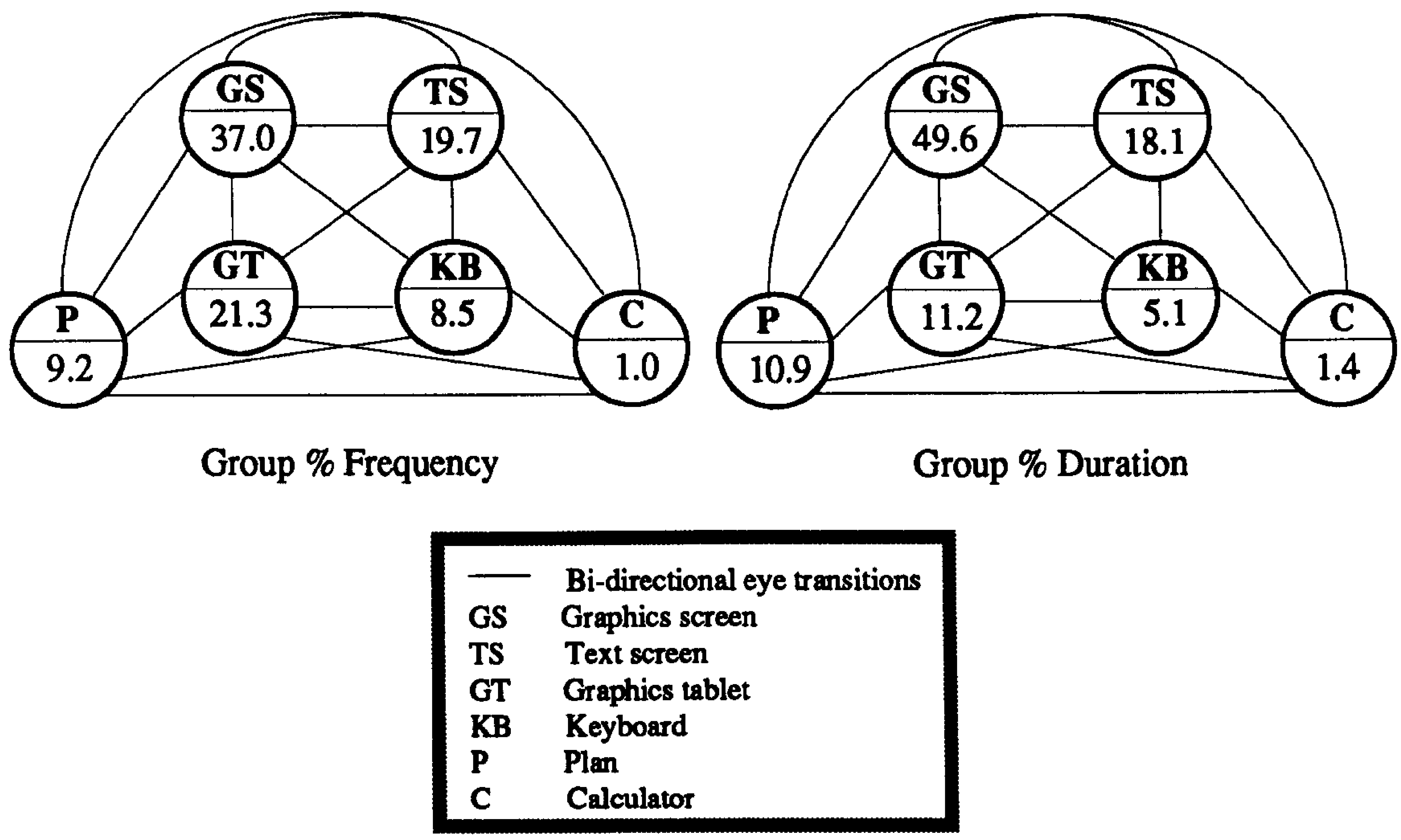


Figure 6.1. Visual behaviour types - Eye gaze to specific targets of interest (n=14)

Studying and checking the plan as work progresses occupied 10.9% of their worktime. The frequency of looking at the plan is 9.2%. Comparatively, the calculator is gazed at less frequently, only 1%. The time spent looking at it takes only 1.4% of the worktime. This further suggests that attention is divided between the task tools, with the plan receiving much more attention than the calculator.

To summarise, the distribution of attention between different parts of the system is not equal. The possible explanations for this will be made in Section 6.4. An important finding is that designers spent 49.6% of their observed worktime watching the graphics screen. Although the combined screen results are much higher (67.7%), this behaviour is still considered here to be non-optimal.

Eye transitions between I/O devices

The results here are aimed at mapping the pattern of eye transitions between input and output devices. There were 315 eye transitions, on average, for a mean duration of 880 seconds, that is, an eye transition from one target to another every 2.8 seconds. Figure 6.2 shows the pattern of eye transitions between input and output devices.

It is evident that many more eye transitions occur from screens to tablet (19.9%) and tablet to screens (20.2%), relative to transitions from the screens to the keyboard (6.6%) and keyboard to screens (7.5%). This result supports the above claim that attention is divided between displays and input devices, with relatively more off-screen eye transitions to the tablet than to the keyboard.

Taking the results of off-screen eye transitions for individual screens, the pattern of eye transitions is as follows: the frequency of transitions from the graphics screen and text screen to the tablet is 14.8% and 5.2%, respectively; the frequency of eye transitions from the graphics and text screens to the keyboard is 5.3% and 1.5%, respectively. This implies that eye transitions to the input devices are contingent on the information being displayed on each screen, and in this case, the graphics screen determines the direction of eye movements to the input devices. Frequent occurrence of eye movements to the input devices is considered here to be non-optimal behaviour.

Hand manipulation of input devices

The group hand manipulation data are presented in Figure 6.3. The results are aimed at identifying the nature of hand activity and the extent of its 'preoccupation' during task performance. (It should be noted that four of the subjects are left-handed, hence for these subjects, the left hand is used to operate the tablet.) From Figure 6.3, it is apparent that:

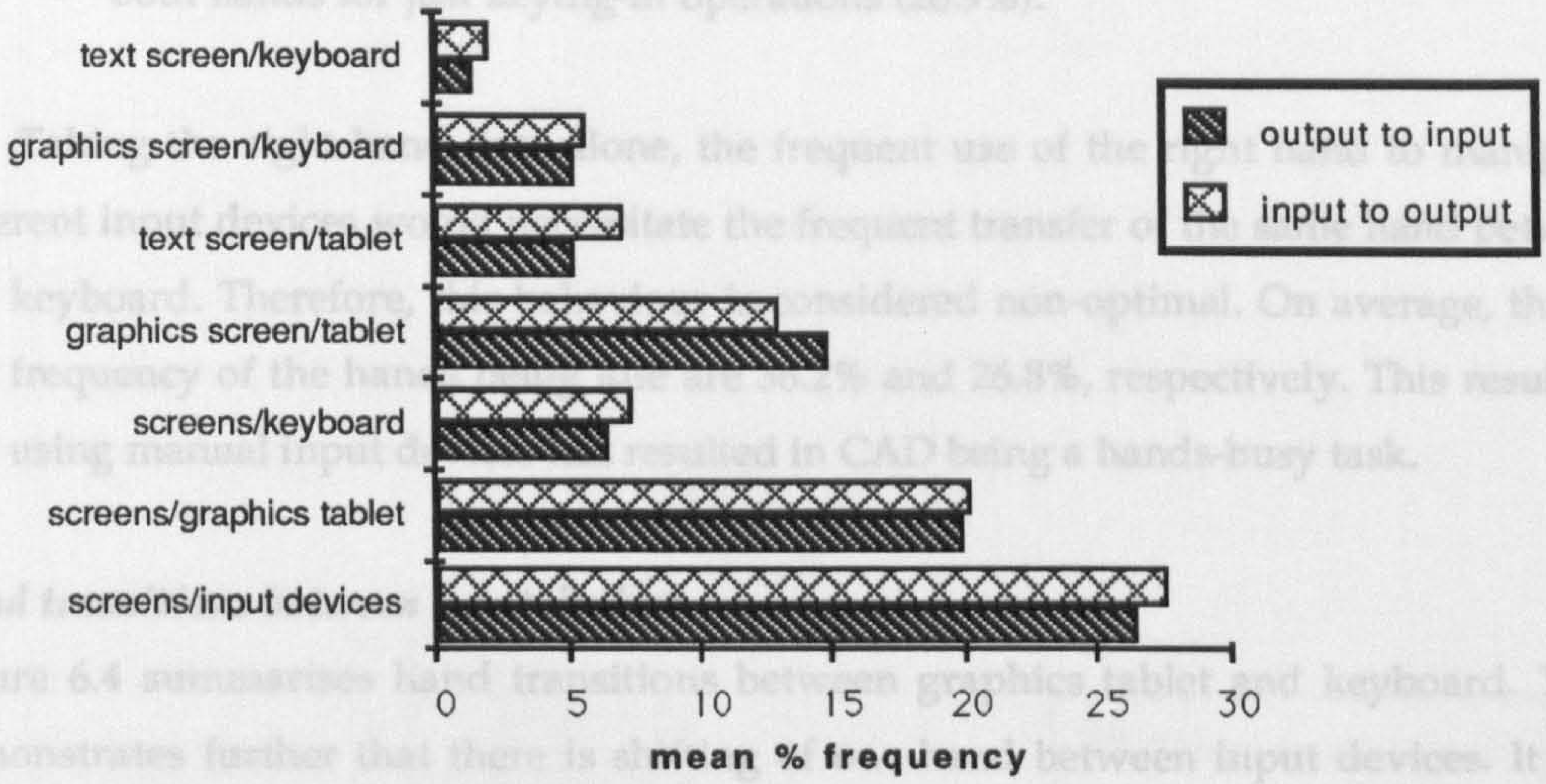


Figure 6.2. Eye transitions between input-output devices (n=14)

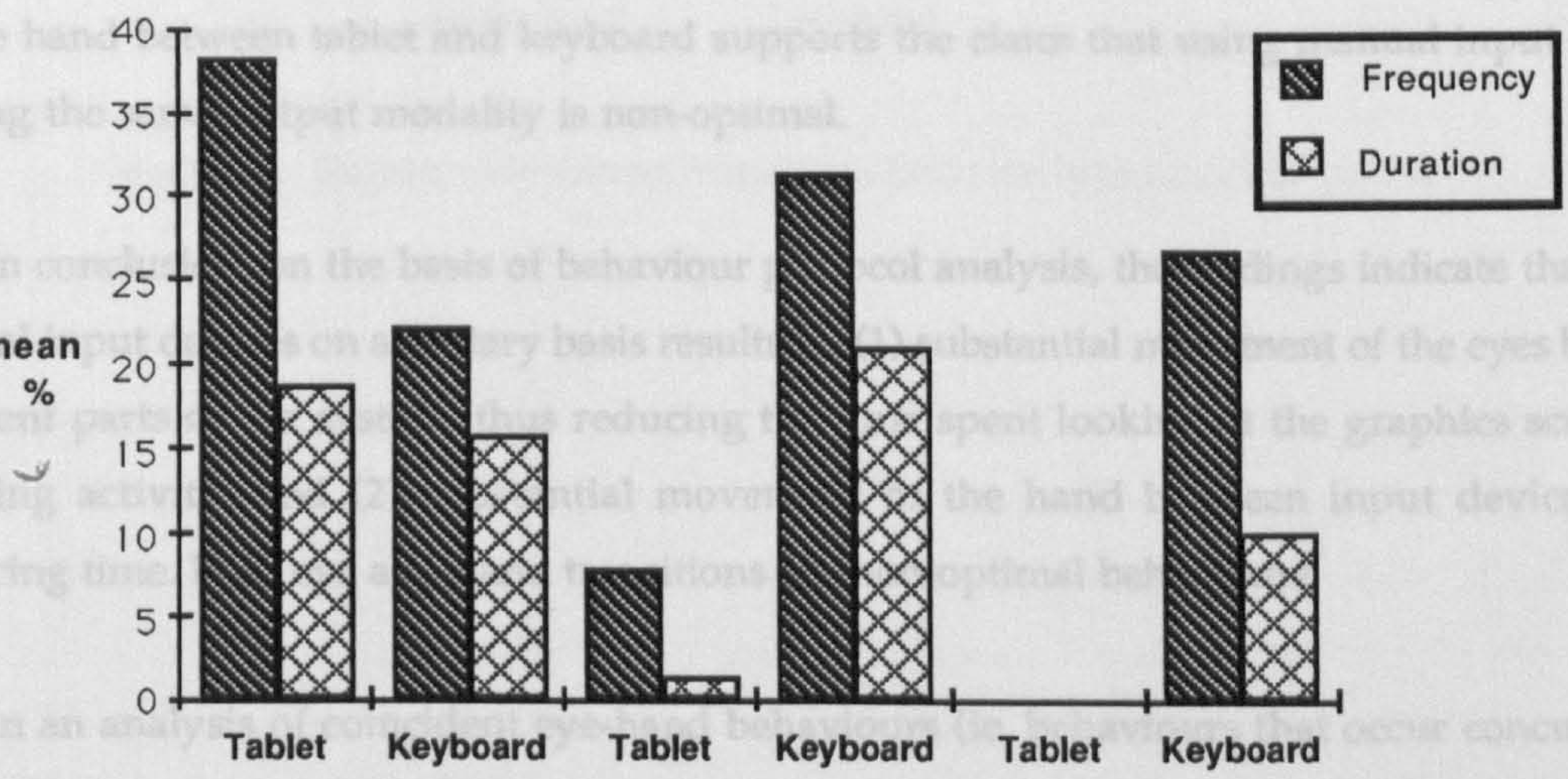


Figure 6.3. Hand manipulation of input devices (n=14)

- the right hand is used to manipulate the graphics tablet (38.3%), except for the left-handed subjects (7.9%), and to key-in information via the keyboard (22.3%);
- the left hand is also used for manipulating the keyboard (31.4%); and
- both hands for just keying-in operations (26.9%).

Taking the right-hand data alone, the frequent use of the right hand to manipulate two different input devices would necessitate the frequent transfer of the same hand between tablet and keyboard. Therefore, this behaviour is considered non-optimal. On average, the duration and frequency of the hands being idle are 36.2% and 26.8%, respectively. This result suggests that using manual input devices has resulted in CAD being a hands-busy task.

Hand transitions between input devices

Figure 6.4 summarises hand transitions between graphics tablet and keyboard. This result demonstrates further that there is shifting of one hand between input devices. It should be pointed out that the results represent the activities of the right and left hands of 14 subjects, independent of their hand dominance or handedness.

Taking the results of the right hand activity, the frequency of right hand transiting to the keyboard from the tablet is 19.5% and vice versa, 19.3%, suggesting that the right hand returns sequentially to the tablet after transiting to the keyboard. The pattern of hand transition is slightly different for the left hand. The left hand transits between tablet to keyboard as frequently as 8.2%, and from the keyboard to tablet, 7.4%. This frequent movement of one hand between tablet and keyboard supports the claim that using manual input devices sharing the same output modality is non-optimal.

In conclusion, on the basis of behaviour protocol analysis, the findings indicate that using manual input devices on a unitary basis results in: (1) substantial movement of the eyes between different parts of the system, thus reducing the time spent looking at the graphics screen for drawing activity; and (2) substantial movement of the hand between input devices, thus incurring time. Both eye and hand transitions are non-optimal behaviours.

In an analysis of coincident eye-hand behaviours (ie. behaviours that occur concurrently) reported elsewhere (Mohd Khalid, 1988), hand transition to input devices (particularly the tablet) can occur 'blindly', that is, without visual monitoring. In other words, users can continue to attend to the screens while reaching for the device. The importance of monitoring the screens visually implies that the CAD task is attention-demanding and any reduction in on-screen gazing or division of attention between display and input devices may incur performance costs to the user.

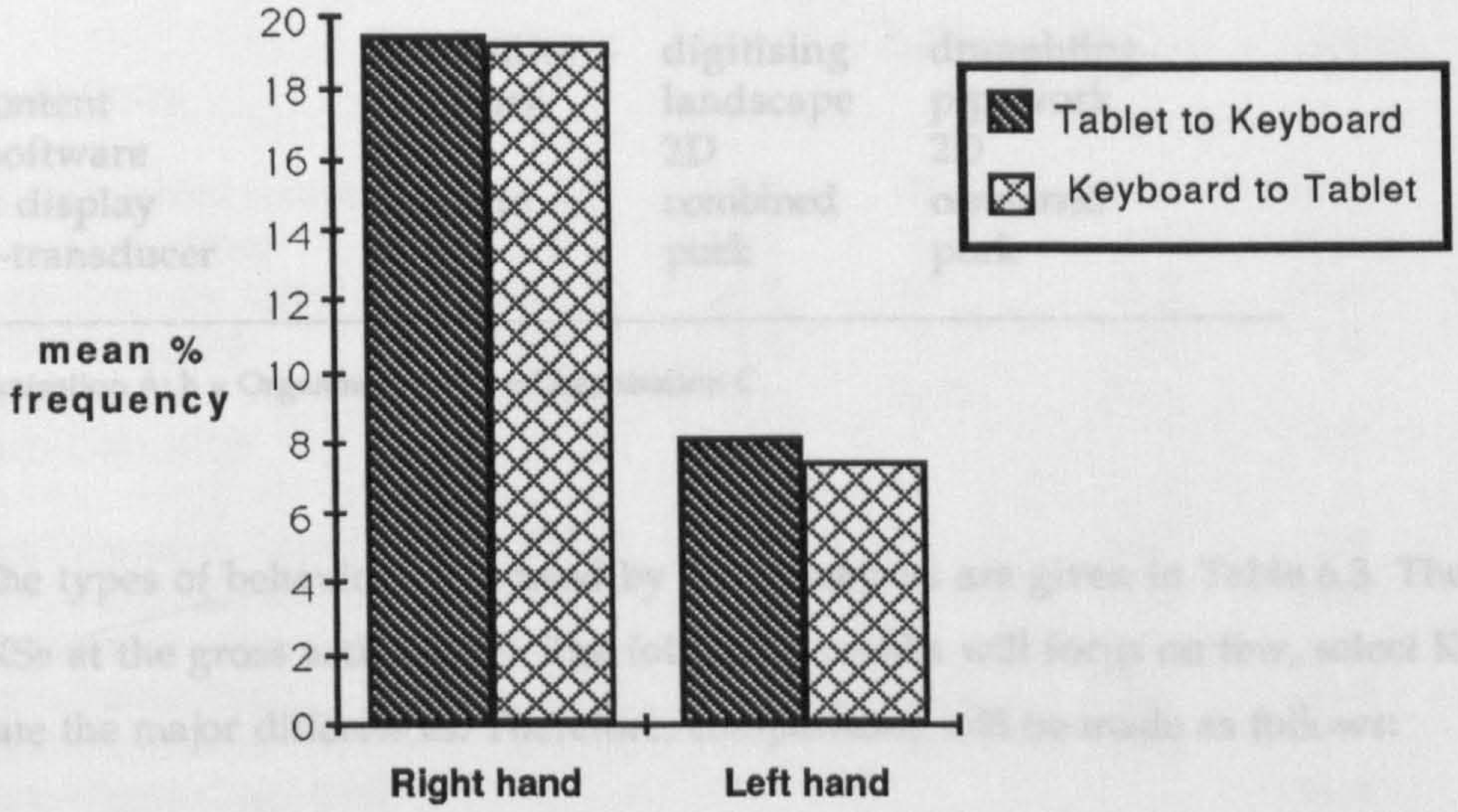


Figure 6.4. Right and left hand transitions between input devices (n=14)

6.3.2 Characterising CAD task knowledge: comparisons of individual data

To compare the recruitment of resources as a function of task and devices, the data from three designers will be used, one from each organisation. To distinguish between data sets throughout the thesis, each subject will be given an identity number, for example, 01S4 means Observational Study 1 Subject 4. A summary of system differences and/or similarities which influenced subject-selection is given in Table 6.2.

TABLE 6.2

A summary of system characteristics determining subject selection

System features	*Subject identity*		
	O1S1[a]	O1S4[b]	O1S8[c]
Task	draughting	digitising	draughting
Task content	pipework	landscape	pipework
CAD software	2D	2D	2D
Screen display	separate	combined	combined
Tablet-transducer	stylus	puck	puck

a = Organisation A; b = Organisation B; c = Organisation C

The types of behaviour KSs used by these subjects are given in Table 6.3. These are low-level KSs at the gross action level. The following results will focus on few, select KSs that will illustrate the major differences. Therefore, comparisons will be made as follows:

Different task, same system: O1S4 versus O1S8

With reference to Table 6.3, O1S4 made an eye transition every 5.0 seconds, compared with O1S8 who made twice as many eye transitions - a transition, on average, every 2.5 seconds. In terms of hand transitions, O1S4 generated fewer transitions (every 24.3 seconds) than O1S8 (every 14.6 seconds). The apparent differences are due, in part, to the type of task performed.

O1S4 performed a digitising task which involved tracing over detail object entities (eg. terrain, rivers, etc.) from a topographical map (ie. plan). This operation requires concentration in order to reproduce an exact representation of the objects on screen. Therefore, the use of the plan is crucial to task performance unlike O1S8 who did not use the plan at all. O1S8, on the other hand, performed a draughting task which involved modifying entities in a stored drawing, hence the need to refer to an offline plan is less vital. In order to perform the task, O1S8 used a calculator to assist in various numerical calculations of the object dimensions. To key-in the information, O1S8 used the keyboard more frequently and for longer periods of the time than O1S4. That is, the frequency of the right hand transiting to the keyboard is 34.6%

TABLE 6.3
Behaviour KSs of three CAD experts

	01S1[a]		01S4[b]		01S8[c]	
Session duration (secs.)	925.2		1314.6		732.1	
Total Eye transitions	340.0		261.0		298.0	
An Eye transition (sec)	2.7		5.0		2.5	
Total Hand transitions	91.0		54.0		50.0	
A Hand transition (sec)	10.2		24.3		14.6	
	%Freq	%Dur	**%Freq**	%Dur	**%Freq**	%Dur
1. **Gross Actions**						
Eyes-gaze-Graphics screen	**29.1**	48.8	**45.6**	56.5	**40.3**	55.5
Eyes-gaze-Text screen	**22.9**	12.2	**7.7**	3.1	**17.1**	14.1
Eyes-gaze-Graphics tablet	**20.0**	12.1	**21.1**	5.5	**23.5**	14.6
Eyes-gaze-Keyboard	**10.6**	7.2	**2.3**	1.2	**7.4**	6.1
Eyes-gaze-Plan	**16.2**	19.2	**22.6**	33.1		
Eyes-gaze-Calculator					**1.7**	1.4
Right hand-transit-G.Tablet	**1.6**	3.1	**34.7**	35.7	**3.9**	2.7
Right hand-transit-Keyboard	**18.0**	8.6	**10.2**	9.3	**34.6**	28.5
Right hand-transit-Stylus	**27.9**	4.8				
Right hand-transit-Puck			**14.3**	6.9	**38.5**	34.5
Right hand-transit-Plan	**9.8**	2.4	**36.7**	41.5		
Right hand-transit-Calculator					**11.5**	29.8
Right hand-idle-G.Tablet	**28.4**	44.6	**38.9**	29.8	**19.5**	25.1
Right hand-select-Data	**25.8**	19.6	**46.0**	50.5	**32.7**	47.9
Right hand-depress-Key	**12.7**	15.1	**9.6**	18.9	**16.8**	13.9
Right hand-calculate-Number					**1.8**	1.9
Left hand-transit-Keyboard	**51.9**	44.8	**100.0**	100.0	**13.6**	1.7
Left hand-transit-Puck					**31.8**	2.3
Left hand-transit-Plan	**25.9**	17.3				
Left hand-transit-Calculator					**9.1**	21.8
Left hand-idle-G.Tablet	**4.8**	5.2	**100.0**	100.0		
Left hand-depress-Key	**88.1**	88.3			**100.0**	100.0

Key:

%Freq = % frequency
%Dur = % duration

a = Organisation A; b = Organisation B; c = Organisation C for O1S8, and 10.2% for O1S4.

Despite using similar devices, the frequency of the dominant hand transiting to the puck differed between subjects. The frequency of O1S8's right hand transiting to the puck is 38.5%, whereas for O1S4 it occurred 14.3% of the time. Again, this is due to the task; in digitising, O1S4 holds the puck when moving from the plan to the tablet, thus the puck is not abandoned, except when the hand shifts to the keyboard which occurred less frequently. For O1S8, the puck is completely abandoned when the hand transits to the keyboard for keying-in operations.

Another important difference is the frequency of the right hand transiting to the graphics tablet. O1S4 made frequent hand transitions to the tablet (34.7%) in order to enter the data points, while O1S8 made less frequent (3.9%) right hand transitions for entering commands. The above demonstrates that the recruitment of resources varies between tasks even when using the same devices.

Different task, different devices : O1S1 versus O1S4

This comparison is to illustrate further that different system configurations will generate different types of behaviour KSs. From Table 6.3, O1S1 made, on average, an eye transition every 2.7 seconds, while O1S4 an eye transition every 5.0 seconds. In terms of hand transitions, O1S1 made a hand transition every 10.2 seconds compared to 24.3 seconds for O1S4. These differences, as explained above, could be due to the type of task performed.

Unlike O1S8, both O1S1 and O1S4 utilised the plan in carrying out the task. For O1S1, the task involved, to a large extent, creating objects according to a large-scale plan. Therefore, gazing frequently at the plan is inevitable; the time spent looking at the plan however differed. The duration was longer for O1S4 (33.1%) than O1S1 (19.2%). As a result of substantial usage of the plan by O1S4, there was also frequent movement of the right hand to the plan (36.7%), compared to 9.8% by O1S1.

The frequency of looking at the graphics screen also differed between subjects. O1S4 gazed as frequently as 45.6% in order to check entity reproduction on screen, while O1S1 gazed only 29.1% of the time on task. The frequency of the right hand transiting to the graphics tablet is more significant. O1S4's right hand transits as frequently as 34.7% while O1S1 just 1.6%. These results further illustrate that the type of task performed and the type of devices used will determine the type of behavioural knowledge recruited.

Same task, different devices: O1S1 versus O1S8

This comparison is to determine whether performing similar tasks (ie. 2D draughting) involving similar applications (ie. pipework) but using different devices will generate different types of behaviour KSs. From Table 6.3, it is evident that S1S1 and S1S8 made similar number of eye transitions (every 2.7 seconds and 2.5 seconds, respectively). There is also little difference in the time taken to generate a hand transition (see Table 6.3).

However, O1S8 used the calculator to assist in numerical calculations while O1S1 did not. The latter instead used the plan which the former did not. This difference in task tool use suggests that although the tasks may be the same, the utilisation of task aids may differ to support performance. To key-in the numerical data, the keyboard was used more frequently and longer by O1S8 than O1S1. The frequency of O1S8's right hand transiting to the keyboard was 34.6%, while for O1S1 it was 18.0%. The frequency of the right hand transiting to the transducer also differed between subjects. O1S1 used a stylus; the frequency of right hand transiting to it was 27.9%. S1S8 used a puck and the frequency was 38.5%. Because the puck had four commonly-used function keys (see Section 6.2.1), this perhaps account for the difference in use.

In conclusion, it could be said that different people have different needs of task tools to assist them in carrying out their tasks. For example, design of printed-circuit boards might require precise mathematical calculations and some systems may not support these facilities, thus users would need to use calculators to do the job offline. Another task aid that might be used differently by users is the CAD manual. To naive users, the manual might be an important learning aid, but might be less needed by experts who have experience with the system. The significance of this comparison is that analysis of knowledge use might reveal some inadequacies or gaps in providing design support facilities to different user groups, suggesting that design of system interfaces (eg. configurability of system components) would require some knowledge of user's behaviour so as to optimise system performance.

6.3.3 Analysis of questionnaire protocol

An analysis of the questionnaire data provided additional information. This information is based on subjects' ratings of their level of skill in using the system and their utilisation of system information in performing the task. The information will be used to verify, in part, the findings of the behaviour protocols. The results will focus on three aspects, as follows:

Subjective ratings of system use

Table 6.4 summarises descriptive analyses of subjects' ratings to Questions 35-39 in the Questionnaire (see Appendix 3) on aspects of system use and performance.

TABLE 6.4
Mean and standard deviations of subjective ratings of system use (n=15)

Categories	*Mean %*	*SD*
Screen looking	75.1	12.8
Device manipulation	50.5	29.8
Tablet use fluency	80.6	15.8
Keyboard use skill	57.7	15.8
Performance	81.5	20.6

With reference to Table 6.4, on average, the designers estimated that they spent 75.1% of the time on task attending to the screens and 50.5% operating input devices. Their estimation of screen-watching is not very different from that obtained in the protocol (ie. 67.7%, see Figure 6.1). But their estimation of device-manipulation is considerably higher than that of the protocol analysis (ie. 34.6%, see Figure 6.3), suggesting that the duration for device manipulation could be higher in total performance or that subjects over estimated the use.

The designers rated high on their fluency in using the tablet (80.6%) but moderately high for the keyboard (57.7%). Since keyboard use requires specific typing skills, this finding reflects a general problem often experienced by users who do not acquire the necessary skills. In terms of their satisfaction with the system, the designers rated high (81.5%) generally. As explained by O1S8, "the system used is very good providing excellent quality of work". Other reasons given include experience and familiarity with the system, and the ability to complete the work on time (see Question 39, Appendix 4).

Utilisation of system information

Table 6.5 summarises subjects' responses to Questions 29-33 of the questionnaire (see Appendix 3) but only a few, select responses will be presented here. This information is required to provide, in part, some possible explanations for gazing at the screens frequently as revealed in the behaviour protocol.

The designers identified three classes of errors made during task performance. First, those arising from mis-keying of input on the keyboard (ie. typing errors). Second, erroneous entry of spatial coordinates (ie. drawing errors), and third, mis-selection of command (ie. pointing errors). In order to identify and correct the errors, 60% of the designers looked at the screens for feedback, while 26.7% were confident of their reentries and found no real need to do so.

TABLE 6.5

Utilisation of System Information during Task Performance (n=15)

Item	Response categories							
	Yes		**Sometimes**		**No**		**No answer**	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Look at command entry via:								
• keyboard	11	73.3	3	20.0	-	-	1	6.7
• tablet menu	10	66.7	5	33.3	-	-	-	-
• screen menu	6	40.0	5	33.3	-	-	4	26.7
Look at screen for:								
• error recognition	14	93.3	-	-	-	-	1	6.7
• error correction	9	60.0	1	6.7	4	26.7	1	6.7
Read prompts?	7	46.7	8	53.3	-	-	-	-
Know prompts without looking?	3	20.0	9	60.0	2	13.3	1	6.7

As with error information, 46.7% of the designers read the prompts at all times while others only sometimes (53.3%). Some (20%) knew the prompts without looking because of their familiarity with the string sequence and command input. 60% of the designers sometimes knew the prompts without directly looking at them. This is because they could detect the prompts in their peripheral vision (see Appendix 4), hence there is less demand for direct looking and processing. The role of peripheral vision in the performance of centrally-demanding tasks is well established (eg. Mohd Khalid, 1981; 1985).

To enter commands via the tablet, 66.7% of the experts always looked at the selection, the rest (33.3%) sometimes monitored the hand when selecting the menu item. However, for command entry via the keyboard, 73.3% tracked visually their finger-pressing activity, while 20% sometimes do this, suggesting that the ability to touch-type does not prevail among these designers, which supports the above finding on keyboard skill rating. Those who could not touch-type tended to rely on auditory feedback for erroneous command input. But in most cases, they looked at the screens for visual feedback, thus resulting in frequent eye transitions between input and output devices.

Device-functionality preference

As described in Chapters 2 and 3, input devices have different functionality. In general, 79% of the designers in this study preferred the tablet for entering commands, drawing and digitising, and 21% preferred the keyboard. The reasons given for either preference were speed, ease of use and convenience. As for alphanumeric data entry, all subjects agreed that the keyboard is best suited for this function. Some of the reasons given for this were that it is quicker, more accurate and the only input method for such data entry.

In conclusion, this questionnaire findings provided support to the protocol findings that the unitary use of manual input devices divides attention between watching screens and manipulating input devices. The off-screen eye transitions, as revealed from the written protocol, are partially due to the need to look at the menu-selection process, and to aid the entry of alphanumeric data using the keyboard. Given that the devices have been optimised to perform clearly delineated functions, their frequent use is thereby increased, resulting in frequent hand movements between devices.

6.4 DISCUSSION

The central issues arising from this study will be discussed in three parts. The first part will review the data in terms of the model, based on the three subjects' data. The second part will focus on the problems documented, and the third part identifies a possible solution to the problem.

6.4.1 Comparisons between individual models of system behaviour

The model was constructed following the procedures described in Chapter 5. Each behaviour protocol was analysed for the sorts of behaviour presented in Table 6.1. These are then categorised into Task Specific and Tool Management KSs. Behaviour types that are concerned with generating or directing the use of solution elements during performance were identified and allocated to the blackboard levels. As explained in Chapter 5, there are essentially three levels (STAM abstraction levels). Therefore, in terms of structure, all three models have similar architecture. The main differences are in terms of the model's content and operation. The content refers to the identity of the KSs recruited; the operation refers to the triggering of particular knowledge types by the knowledge executor.

Comparing between individual models showed that O1S4 recruited more graphical data at the SubTask level of the blackboard than O1S1 or O1S8. This is because the task involves digitising coordinates from a topographical map. The greater recruitment of Eyes and Hand KSs to aid the digitising process implies that the allocation of resources is determined by the type of task operation being performed. In other words, the scheduler applies different control algorithms (eg. duration of KS use) to enable task execution. O1S1 and O1S8 performed similar draughting tasks and tended to use a variety of Tool Management KSs to perform the tasks. But the behaviour types differed between them: O1S8 recruited more Keyboard KSs at the Movement level than O1S1. Also, O1S8 recruited Calculator KS while O1S1 recruited Plan KS at the Action level. This, then, indicates that the recruitment of knowledge depends on task requirements, and that different KSs will be recruited at different levels. How the knowledge is applied is managed by the scheduler. The latter uses a number of criteria, such as the user's skill in manipulating the tools (eg. keyboard skill), the importance of the KS to support performance (eg. the use of a plan to increase design accuracy) and the history of KS use (eg. the frequency of Tablet KS use).

In short, the model provides a means of understanding how knowledge is recruited during CAD performance, and indirectly suggests the triggers for particular behaviours, which could be used as a basis for optimising design behaviour and task performance. In particular, the model has helped to identify critical behaviours that are relevant to the task. The next section reviews particular behaviours and their relationship with user skill.

6.4.2 Problems documented

There are two central problems identified in this study. First, the problem of having to monitor visually the input devices during manipulation because of inadequate skill (ie. typing and tablet-selection skills). This neccessitates the transiting of eyes from the screens, in particular the graphics screen, to the input devices. Second, the problem of shifting hands between input devices sharing the same manual mode (as discussed in Chapter 1 and predicted in Chapter 4).

In CAD, as explained in Chapter 4, the screen(s) replace the traditional drawing board; therefore, in most circumstances vision should be maintained on the screen(s) for a greater proportion of the time on task. In this study, this happens only 68% of the observed worktime (see Figure 6.1). The remaining proportion of the time is spent mostly operating input devices and handling task tools.

The findings support the claims made by: (a) Van der Heiden and Grandjean (1984) that CAD operators spent between 48 and 68 per cent of their worktime at the CAD terminal watching the screen; 26-48 per cent operating the tablet; and 14-24 per cent operating the keyboard; and (b) Monk (1986) that shifting between input devices sharing the same modality incurs cost.

There are two possible explanations for the frequent off-screen eye transitions to the input devices. Firstly, the performance of CAD task requires the entry of graphical data, which in turn is contingent on command entry. Because not all designers are skilled at operating the tablet without visual aid, the need to gaze away from the screen to the tablet for this purpose necessitates considerable head movements. Secondly, the CAD task also requires the entry of alphanumeric data for specifying object parameters and annotating the drawing. This entry is usually made via the keyboard. Because not all the designers can touch-type, the need to monitor visually the keying-in operations results in further eye transitions off-screen.

Associated with the above problem is the use and transfer of hands between input devices. In terms of hand use, users spent 67.1% of their worktime operating the devices manually (see Figure 6.3) for drawing and entering commands as well as text. Of this proportion, about 30% of the hand activity was aided visually, in particular the entry of command and text. The remaining proportion of the operation involves drawing which does not require visual attention to the hand. This substantial deployment of resources for manual operation keeps the hands busy for a significant proportion of the time on task. Thus, the notion that CAD is a 'hands-busy' task is confirmed in this study, as revealed from the hand-idleness data.

The frequent transfer of hands between input devices incurred some costs to the user. For certain users (eg. O1S8), the time taken to use the puck after transiting from another input device (eg. keyboard) seems longer than another user (eg. O1S4). Because of individual differences in device manipulation skill, the frequent switching between devices sharing the same modality may impose additional demands on the user and might slow performance on all aspects of the task. Karat et al. (1984) illustrate the costs which can be involved in shifting between keyboard and pointing device (eg. mouse). Thus, there is reason to believe that using manual input devices on a unitary basis may have a detrimental effect on task performance in general.

In addition to the reasons given above, gazing off-screen has potential for causing decision-type interference to occur in task execution. This is because designers make decisions about the many parameters of the various parts of the object under design (see Chapter 3). Therefore, the need to gaze away frequently might disrupt the current state of the drawing. Secondly, as revealed from the questionnaire data, screen-watching also serves to elicit task-relevant information (eg. error messages, prompts), which serve as feedforward and feedback mechanisms. In the absence of this information, guessing what the messages might be could easily lead to errors, which in turn could lead to more reentries, and consequently increased off-screen transitions. In short, the performance costs incurred by unitary use of manual input devices can be substantial.

6.4.3 Possible solution

It is possible that the problems of eye and hand transitions outlined above may be alleviated by the use of a speech recognition system. The potential of unitary speech input as a solution to manual input will be discussed in detail in Chapter 8. Suffice it to say that with speech replacing some of the functions performed by manual input, the problems of eye and hand transitions as documented here may be resolved. This is because if speech input is used to enter commands, the user will be able to gaze at the screen(s) while speaking, and the hand(s) will only be used to manipulate the tablet for drawing, whilst the keyboard for entering text. This should result in fewer off-screen and between-device transitions.

6.4.4 Conclusion

The CAD systems in this observational study employ a graphics tablet and an alphanumeric keyboard as input devices. This combination places considerable demands on the user, who has to look away a considerable proportion of the total worktime at the terminal in order to locate target commands on the tablet menu or to locate target alphanumeric and/or function keys on the keyboard. In addition, the user has to shift one hand frequently from the tablet to the keyboard in order to perform the above functions. This division of attention between the screens and input devices, and the shifting of hand(s) between input devices may have adverse consequences for task performance.

The effect of non-optimal behaviour on performance could not be ascertained from this exploratory study due to: (1) the variability in the task content and the devices used; and (2) the fact that this study was not set out to measure performance. This, then, requires further investigation into the use of manual input and a comparison with unitary speech input - a proposed solution to the problems concerning manual input.

6.5 SUMMARY

This observational study of CAD experts at work, aimed at documenting the nature of non-optimal behaviour in using unitary manual input to perform CAD tasks, has shown two things. First, the use results in considerable off-screen eye transitions. Second, the use incurs between-device hand transitions. Both types of behaviours are non-optimal and may incur performance costs to the user. Given that this study was not designed to measure performance, the relationship between non-optimal behaviour and performance could not be established. An alternative to manual input is the use of unitary speech. Its potential as a solution will be the subject of subsequent investigations.

Data from this exploratory study were also used to develop a blackboard model of system behaviour based on a characterisation of visual and manual behaviours of these experts. The group model was used to demonstrate the problems of non-optimal behaviour; the individual models were used to illustrate the differences in knowledge recruitment due to the type of task performed and the type of system in use. Subjects' written protocols gathered by means of a questionnaire provided the information to substantiate the observations made.

NEXT CHAPTER HIGHLIGHTS

Chapter 7 describes a study aimed at optimising the performance of a demonstrator system to be used in subsequent experimental investigations of using speech input as a possible solution to the problems of manual input. This optimization is considered essential in order to provide an understanding of the performance characteristics of each input device prior to experimental usage. Factors determining the selection of various system components are also discussed.

CHAPTER 7

Study 2: Optimising the Performance of a Demonstrator System

Overview

7.1 Introduction

7.1.1 Why optimise?

7.2 Description of demonstrator system

7.2.1 Speech recogniser: PRONOUNCE

7.2.2 Graphics tablet: SUMMASKETCH

7.2.3 AutoCAD software

7.3 Optimisation method

7.3.1 Subjects and task

7.3.2 Procedure

7.4 Analysis and discussion

7.4.1 Human factors criteria

7.4.2 Application task: measures of system performance

7.4.3 Questionnaire findings

7.5 Device assessment and recommendations

7.5.1 Speech recogniser

7.5.2 Graphics tablet

7.5.3 Other system aspects

7.5.4 Conclusion

7.6 Summary

Next chapter highlights

CHAPTER 7

Study 2: Optimising the Performance of a Demonstrator System

OVERVIEW

This chapter describes an optimisation study aimed at understanding the performance characteristics of the speech recogniser and the graphics tablet in order to ensure the best configuration for a demonstrator CAD system. Failure to carry out an appropriate optimisation, prior to the experimental investigation, may render interpretation and generalisation of the experimental results difficult. The observational study described in the previous chapter provides the basis for specifying the requirements of this system.

The procedure involves optimising system performance in relation to: (1) technical recommendations concerning the use of the input devices; (2) human factors criteria derived from the literature; and (3) an application task, that is, a CAD draughting task. On the basis of descriptive assessment, the strengths and limitations of each device were documented, and recommendations about configurations were made to improve device use. This led to specific modifications of the experimental system.

7.1 INTRODUCTION

7.1.1 Why optimise?

The evaluation of any set of devices requires, as a precondition, the optimisation of each to ensure that an appropriate comparison or series of comparisons can be made (Long & Johnson, 1982). Failure to optimise a device may lead to the evaluation of a device that is not representative of its own generic class. In the absence of optimisation, gross differences in performance between devices are likely to be captured but it may not be possible to ascertain what proportion of the differences are due to characteristics of the device itself. For example, a stylus needs to be held upright, unlike holding a pen (see Chapter 2), to ensure optimal performance. Failure to understand this behaviour and its possible effect on performance can lead to erroneous conclusions.

There are two main goals of this optimisation study. The first is to provide an understanding of the performance characteristics of the speech recogniser and graphics tablet prior to their experimental use. Given that the selection of these devices is based on the fact that each is a leading example (ie. popular, widely-used, etc.) of its own class of input devices, this optimisation is essential to ensure that they are both *representative* of their input device class. This, then, enables comparisons to be made between typical examples of input devices.

The second and equally important goal is to understand the performance of the demonstrator system to be used in the subsequent research. This means optimising the potential system in terms of certain human factors requirements. As defined in Chapter 4, a demonstrator system is a CAD system that demonstrates what the best configuration of I/O devices should be. The next section describes this system and discusses the factors governing the selection of the various system components.

7.2 DESCRIPTION OF DEMONSTRATOR SYSTEM

The observational study (Chapter 6) suggests that a common configuration of CAD systems utilises two screens with combined graphics and text on each screen. Such a combined dual-screen configuration may make it difficult to assess the relationship between screen-looking (ie. behaviour) and the utilisation of information during performance. Thus, this demonstrator system will employ a separate screen for graphics and text. It should be pointed out that this configuration does occur in real-world systems (see Chapter 6), although it may not be the most common one. This clear-cut separation of graphical information from text will enable better assessment of the utilisation of information during CAD performance.

Additionally, it should be noted that the configured system is representative of microcomputer CAD systems rather than specialised mainframe or minicomputer systems. The hardware and software chosen for this system met several basic requirements. Time and budget restrictions required the equipment to be relatively inexpensive, readily available, and to be compatible with existing equipment within the research environment. Within the various constraints, the chosen hardware which made up the demonstrator system was the following:

- an IBM PC-XT computer with 10Mb hard disk, and a standard alphanumeric keyboard with QWERTY layout;
- two terminals; one a colour monitor for graphics display, another a monochrome monitor for text display;
- a SUMMASKETCH graphics tablet with a single-button stylus as the transducer;
- a speech recognition system called PRONOUNCE, fitted with a SHURE SM10 headset, noise-cancelling microphone;
- a Hewlett-Packard plotter for producing a hard copy of the drawing output; and
- an IBM PC graphics printer for generating a hard copy of speech input.

The application software was AutoCAD release 2.17.

The choice of an IBM personal computer (PC) is based on the fact that it has been a de facto standard for micro-computers. As such it provides support for a variety of computer peripherals and application programs. This IBM system runs the MS/PC-DOS operating system. Despite the constraints relating to DOS (see First Draft, July, 1989), the existing base of DOS users has increased to more than 15 million (source: PC Week, April 4, 1989). Given the large user base, IBM itself will continue to support DOS development (Utley, 1988).

The IBM PC-XT version, for example, can form the basis of an efficient CAD workstation at a cost of around £5,000. For this reason, the use of PCs, as opposed to mainframes, is commonplace particularly in companies with small-scale economies. In addition to CAD, many speech recognition systems are also designed for the IBM machine, in particular the PC-XT/AT range. This thus favours the use of an IBM PC-XT computer in this investigation.

The IBM PC-XT comes with a standard alphanumeric keyboard, including ten function keys and four cursor keys. The graphics terminal employs a CGA (colour graphics adaptor) which enables four colours to be displayed at a time, from a choice of 16 colours, with a resolution of 640 x 200 pixels (or picture elements). The alphanumeric text terminal employs an MDA (monochrome display adaptor) which displays text in green.

Having described the processor system, the next section will describe the input devices and CAD software in terms of their characteristic features and the factors influencing their selection.

7.2.1 Speech recogniser: PRONOUNCE

In Chapter 2, some of the human factors issues in speech recognition were discussed. Within these considerations, the choice of a speech recogniser would be one that: has many essential features like good-size vocabulary, easy training, etc; can be used in most applications like word processing, database management, besides CAD; can be configured with the rest of the equipment; is reasonably priced; and is portable, though this is not crucial. Appendix 6 gives a summary of system features on which a few, select speech input systems were compared.

On the basis of the above criteria, PRONOUNCE version 1.20 by Microphonics Technology Corporation was selected for use in this study. PRONOUNCE satisfied the following:

- it works on IBM PC-XT/AT and all IBM workalikes with at least 256K and DOS 2.0 or above; thus it interfaces through MS/PC-DOS with most application software, including CAD;
- it requires an input between 500 milliseconds to 2 seconds in duration; as such it can accommodate phrases as well as single words;

- it has an online vocabulary size of 256 words or phrases and each word/phrase may generate up to 255 keystrokes;
- it comes with a headset, background-discriminating microphone which allows it to be used in noisy environments;
- it is reasonably-priced costing less than £800.

The recogniser consisted of a full-length circuit board and its operation required 68K of the computer's RAM. A floppy disk with PRONOUNCE programs and sample vocabularies are provided with the system, as well as a User Manual (Microphonics, 1984). The recogniser can be turned on and off only with a speech command, which may pose a constraint to users. Some systems have alternative ways (eg. a scroll lock key on the keyboard) of activating the recogniser, besides spoken commands (see Martin, 1989). This facility is useful as a backup to verbalised input.

Besides satisfying the necessary requirements, PRONOUNCE was one of the few commercially available systems that offerred a *single pass training*. By this is meant that each word need be trained just once for the system to remember and recognise it afterwards. Other systems such as VOICESCRIBE and IBM Board (see Appendix 6) would require more than one verbalization during enrolment in order to achieve comparable recognition accuracy.

Because of this single-pass training feature, PRONOUNCE will only allow the user to vary the speech input by 20-30% during usage. This includes the volume, enunciation, accuracy and length of each phrase or word (see PRONOUNCE User Manual, p. 12). For most connected word recognisers, usage performance tend to range between 50-80% (Chapter 2). (A distinction between speech performance figures given by manufacturers and those obtained during enrolment and usage was raised in Chapter 2.) Following the allowable range for speech variability, speech usage performance of this speaker-dependent, connected word recogniser will be in the region of 70-80%. This makes PRONOUNCE representative of its class of speech recognisers.

7.2.2 Graphics tablet: SUMMASKETCH

As discussed in Chapters 2 and 3, the choice of a graphics tablet depends primarily on its resolution, accuracy, portability (or size) and style. Resolution can range from 100 to 1,000 lines per inch (lpi), and accuracy from 0.025 to 0.001 inch. Of the two features, accuracy is the more important consideration because even the best resolution will not help if the apparent cursor status is not being correctly relayed to the software (Bickel, 1987).

Appendix 7 provides a comparison of available digitiser types. The Summagraphics company is the market leader for digitisers (source: CADCAM International, June 1987). Their

SUMMASKETCH model 1201, consisting of a portable tablet measuring 12" x 12", will be used in this research. SUMMASKETCH has the following features:

- it is designed for the IBM PC and 100% IBM PC compatible systems;
- it is a high-resolution device based on electromagnetic technology; with a resolution of 1000 lpi of active area, and a claimed accuracy of 0.025 inch;
- it supports a one-button stylus as well as three-and four-button pucks which are easily interchangeable; and
- it is a low cost model costing less than £500.

The tablet can be adjusted to varying angles of tilt to suit the user. For details on its technical specifications, see SUMMASKETCH User's Handbook (Summagraphics, 1984).

7.2.3 AutoCAD software

The number of available CAD programs has increased to meet the needs of a rapidly growing user base. Thus, choosing a CAD system is not an easy matter. The popular literature in CAD (eg. CADCAM International, CAD User, PC User, PC Week) provides useful guides to buyers on how to choose a CAD system. Generally, the advice includes to look for a system that: is easy to learn and work with; has an open architecture; is flexible, expandable and compatible with other CAD programs, etc. (CAD User, September 1988, p. 56-60). Appendix 8 compares some PC-based CAD packages, particularly those that support 2D draughting. Of these, AutoCAD is considered to be the best established of the low-cost PC-based 2D CAD packages (Massey, 1988; Lang, 1985).

AutoCAD was developed in the USA by Autodesk Inc. and is claimed to be "the world's most famous CAD package. By CAD standards it has been staggeringly successful, selling over 150,000 copies" (Bright, 1988, p. 45). AutoCAD now has an installed base of over 500,000 users worldwide (source: CAD User, July/August 1989), with a full range of upgrade products, such as AutoSketch, AutoCAD AEC Architectural, AutoSolid and AutoShade. In the UK itself, there are over 15,000 AutoCAD users (source: First Draft, July 1989). AutoCAD is used in teaching at some 230 educational establishments in the UK, from universities and polytechnics to Manpower Services Commission ITECs (Information Technology Centres) (source: AutoCAD Expo Europe Catalogue, 1987, p. 8).

The real strength of Autocad lies in its direct support for a large number of peripherals or add-ons which enhance its capabilities, thus making it more applicable to different types of CAD work. Because of its open nature, dealers like it because they can adapt the package to make more money and end users like it because it can be tailored to their requirements (Bright, 1988).

AutoCAD has changed a great deal since it was first introduced at the end of 1982. The software has evolved from a basic 2D draughting system for the IBM PC to a sophisticated programmable 3D drawing package running on hardware ranging from PCs to workstations by SUN, Apollo and DEC. However, "AutoCAD now seems to fall in the gap between 2D and 3D systems. It offers much more than the average 2D drafting system but it doesn't quite make it as a 3D surface modeller" (Bright, 1988, p. 49). So, although Autodesk are working on the 3D side of AutoCAD, 2D will still dominate for some time to come, and will require better applications support and integration (source: CAD User, June, 1989).

Given the problems in learning and using 3D systems (see Chapter 3), new users to CAD are advised to start with 2D first (eg. Lang, 1985; Haase, 1989). Since this study will involve naive users, the use of a 2D draughting system is appropriate. Also, as discussed above, a 2D system is better established than the 3D. To summarise, the choice of AutoCAD is based on a number of factors:

- its popularity and worldwide usage;
- it is designed specifically for PC-based systems;
- it is general-purpose for a wide variety of applications;
- it provides many design tool facilities;
- it is a mid-priced package costing £2,500;
- it is easily customisable; and
- it has a command language that is semantically meaningful.

The latter two features are crucial in the context of this research. The tailorability aspect will enable the system to be easily configured to suit experimental requirements, while the dialogue aspect will promote acceptability and speed up user learning, especially in the case of naive users.

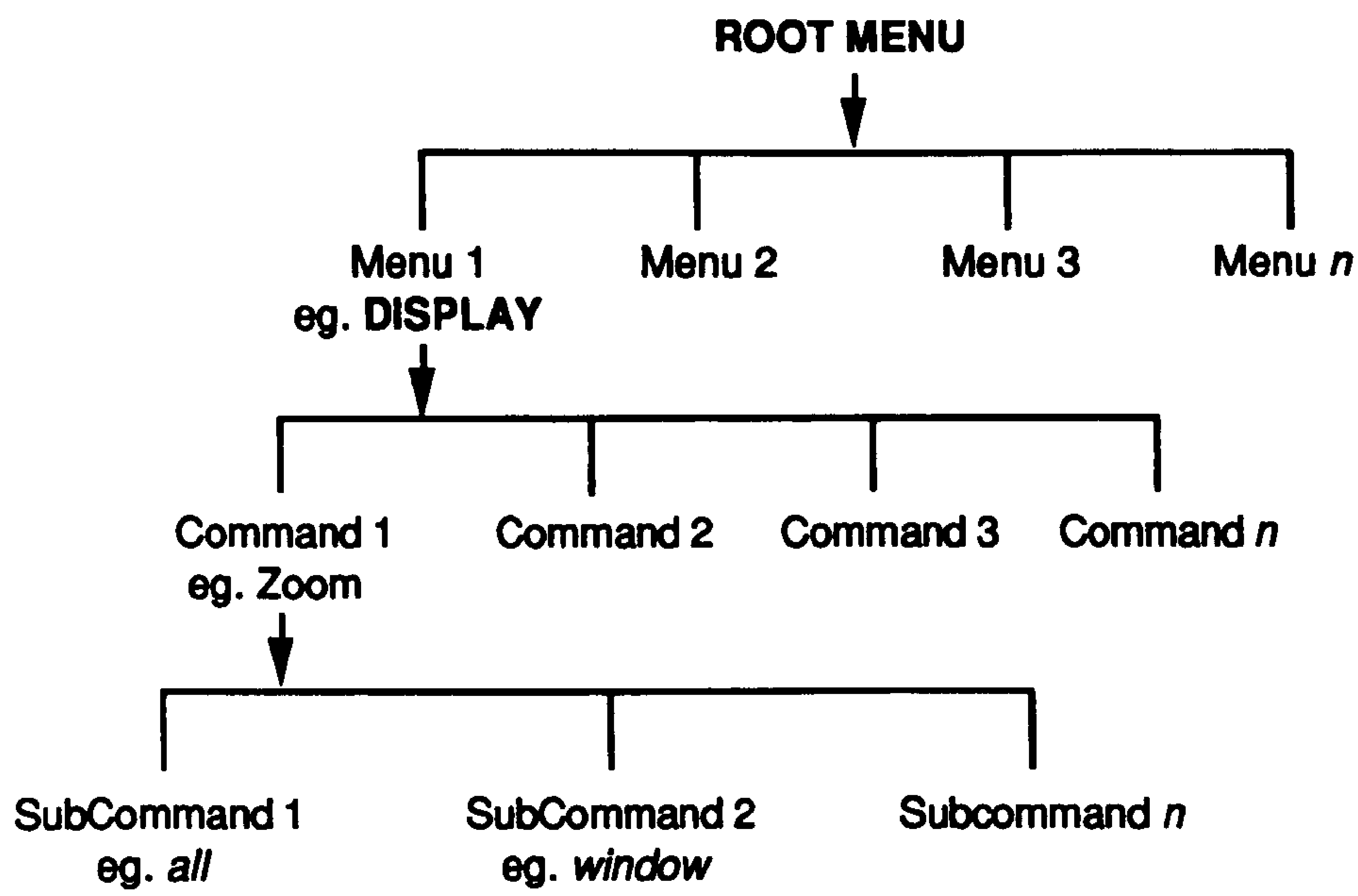
AutoCAD version 2.17

Version 2.17 of AutoCAD offers only 2D drawing on the standard 640 x 200 pixel graphics display. It comes with a User Reference Manual (Autodesk, 1985). Since this manual might be too complex and cumbersome for first-time users, a simplified version was compiled for this study. Appendix 9 provides a brief introduction to the experimental system and a list of AutoCAD command names and their definitions.

Table 7.1 shows the AutoCAD command structure and a summary of major commands, some of which will be used in this research. AutoCAD has a flat command structure as opposed to a heap or hierarchical command structure (see Inside AutoCAD, 1985, p. 284). The former allows one to directly select any command at any time, although the commands are presented hierarchically; the latter only allows one to select a desired command after having selected a

TABLE 7.1

AutoCAD Command Structure, Root Menu and Major Commands



Root menu	Some major commands
BLOCKS	• block:; insert:; base:; attrib; wblock:; attdef:; attdisp:; attedit:; attext:
DIM:	• dim:; linear; angular:; diameter:; radius:; center:; leader:; aligned:; etc
DISPLAY	• pan:; qtext:; redraw:; regen:; view:; zoom:; qtext
DRAW	• arc; circle; insert:; line:; solid:; text:; trace:; point:; rectang:;ellipse:
EDIT	• array:; break:; chamfer:; change:; copy:; erase:; fillet:; move:; repeat:
HATCH:	• hatch:
INQUIRY	• area:; dist:; id:; list:; status
LAYERS	• layer:; linetyp:; ltscale:
MODES	• axis:; blips:; coords:; dragmod:; grid:; ortho:; osnap:; snap:; tablet:
PLOT:	• plot:
UTILITY	• apertur:; files:; help:; limits:; menu:; style:; units:; save:; end:; quit:

number of commands in the command structure. Commands are grouped for convenience, in the form of menus (eg. DRAW, EDIT, DISPLAY, UTILITY, etc.). As explained in Chapter 3, a command performs fundamental interaction (FITs) and/or control task functions.

With reference to Table 7.1, the basic drawing elements in AutoCAD are **lines**, **traces** of any width, **arcs**, **circles**, **solids** and **inserts** of other drawings. Drawings can be annotated with **text** of any size, at any position and angle, from a variety of fonts and styles. Drawing elements can be positioned on the screen by 'freehand' pointing - with or without the use of **grid**, **snap**, **ortho** and other drawing aids - and in most cases they can be interactively '**dragged**' into position. Where exact positions or dimensions are required, these can be entered via the keyboard, expressed either by absolute x, y coordinates, lengths or angles, or as relative x, y displacements or distance and angle from the last point.

Drawing aids such as **object snap (osnap)** lock drawing elements onto reference points on existing objects: **endpoint** of line or arc, **midpoint** of line or arc, **centre** of arc or circle, **intersection** of lines, arcs or circles, **tangent** to an arc or circle. A bi-directional **zoom** facility allows working on drawings at any level of detail. Particular 'windows' can be moved in any direction by the **pan** command.

A set of editing commands allows drawn objects to be **moved**, **copied**, **changed**, **rotated**, **mirrored**, partially or completely **erased**, and **scaled** vertically and horizontally. Any angular corner can be replaced by a **fillet**, or a **chamfer**. Drawings can be created on an unlimited number of **layers**, each with user-defined alphanumeric names. Layers can be displayed and plotted with different non-continuous **linetypes**, and in different **colours** if the system's display and/or plotter allow. Layers can be turned on and off as desired. **Hatching** can be generated automatically at any angle and pitch, either from the library of over 40 standard patterns provided or from user-defined additions.

The **status** command reports on the current state of a number of drawing parameters, such as overall size, size of current window, base point, etc. The **limits** or overall size of a drawing are initially set by default, but may be changed at any time. For details on this package, see AutoCAD User Reference Manual (Autodesk, 1985) and *Inside AutoCAD* by Raker and Rice (1985).

7.3 OPTIMISATION METHOD

This study was conducted in a laboratory setting, in the Ergonomics Unit, University College London. The optimisation method involved both direct and indirect observation. In the case of direct observation, a form was used to record the behaviours observed, thus providing descriptive analysis of behaviour and performance. With indirect observation, the behaviour was recorded on video.

7.3.1 Subjects and Task

This study involved only two CAD-naive subjects, referred to here as O2S1 and O2S2. (Note: O2 means Observational Study 2.) Both subjects were British, bespectacled and right-handers. O2S1 was a female, aged 34 years with little computer experience; whilst O2S2 was a male, aged 22 years with 2 years of computer experience. This diversity in age and experience may account for some of the variability in performance.

The CAD task chosen for this study was one provided by Autodesk in their training of educational AutoCAD users. The task involves designing a stanchion mount (see Appendix 10) with drawing instructions provided as a guide. As such, it was considered a minimal task (ie. well specified and with few task requirements), but representative of CAD draughting tasks. The same task was used in both the training and optimisation phases so as to neutralise any effects of the task. The commands used for the task and the drawing procedure are summarised in Appendix 11.

7.3.2 Procedure

The study was conducted in three phases, as follows:

Phase 1 - *Training in AutoCAD* (1.5 hours).

This phase includes a brief demonstration, practice session and the performance of the above CAD task.

Phase 2 - *Optimising PRONOUNCE and SUMMASKETCH* (2 hours).

This phase involves three sequential activities:

(1) optimise to device recommendations;

(2) optimise to human factors criteria; and

(3) optimise to application task, as above.

These activities were conducted separately for each input device. Activities 1 and 2 were aided with a recording form-cum-checklist (see Appendices 12 for speech recogniser and 13 for graphics tablet versions), and a stopwatch to record user response times to a particular command. Activity 3, however, was recorded on video using the same set-up as in the observational study of CAD experts, described in Chapter 6.

Phase 3 - *Debriefing subjects on their experiences* (30 minutes). This phase involves verbal probes and the use of a short questionnaire comprising both closed and open-ended questions (see Appendix 14).

The following paragraphs elaborate on the optimisation procedure.

Optimise to device recommendations

These are recommendations in the user handbook or manual on how the device should be used. For example, the headset microphone for the recogniser should be positioned to the side of the mouth, NOT in front of it, and there should be a gap of 1.5 inch (or a thumb placed) between the microphone and corner of mouth (see Appendix 12). This positioning is crucial in the optimal use of speech devices, as discussed in Chapter 2. Thus, the purpose of this optimisation is to check that the device is in good working order as prescribed in the manual and to identify problems with respect to its use. These findings will be used to prepare a set of instructions for optimal use of the device, which subjects should comply with in the actual experiments.

Optimise to human factors criteria

The criteria are obtained from a review of the human factors literature concerning input device use in general (eg. Whitefield, 1986a; McCauley, 1984; Ritchie & Turner, 1975; Bailey, 1982). These relate mainly to recognition/selection accuracy, response speed, ease of use and flexibility. Appendices 12 and 13 described how these criteria are measured for the recogniser and tablet, respectively. For example, in the case of the tablet, pointing or selection accuracy is crucial to optimal performance. If the pointing area for each command on the tablet menu is too small, then this could lead to pointing errors (ie. aiming at the wrong command), as identified in Chapter 6. Alternatively, if the tablet menu is too cluttered with information, this might slow visual search of target item, and consequently, task performance. Therefore, the purpose of this optimisation is to assess whether the device meets human factors requirements and to examine the possibility of improving system performance in accordance with these criteria.

Optimise to application task

The purpose of optimising the input device to a CAD task is to examine how the system as a whole performs in the context of a real task. In other words, this finding will help to identify potential problems of the system in performing a CAD task. Assessment of system performance will be made in terms of: (1) time on task (ie. task completion time); (2) product quality (ie. number of errors in the drawing); and (3) production (time) cost (ie. time per drawing entity). This measurement is in line with the methodology outlined in Chapter 4.

7.4 ANALYSIS AND DISCUSSION

This section will present and discuss the principal findings from the optimisation process.

7.4.1 Human factors criteria

The findings concerning each input device will be presented in turn.

Training

The literature identified two types of speech recogniser training: template training of the device, and vocabulary training of the user (Chapter 2). In this study, template training involves two stages. First, setting the voice level by verbalizing the word "Pronounce" twenty times, followed by training a set of five words, specified by the software, four times. This was then checked for recognition by re-verbalizing each word twice in random order. The second stage involves training a 'test' vocabulary of 24 commands (21 AutoCAD commands plus 3 PRONOUNCE commands) by verbalizing it once, then checking for recognition.

User training involves two stages: first, learning how to speak the vocabulary such that the recogniser will work, for example, identifying enunciation problems, speaking duration, volume, etc. Also, subjects are encouraged to retrain a particular word that did not meet a threshold score of 8 in the check mode or has the potential of being confused with another acoustically similar word (eg. Grid and Quit). Secondly, users are trained on how to use the device effectively. For example, how to position the headset microphone during use in order to optimise input reception; how to de-activate the system when drawing so as to reduce task interference, etc. This is termed here device training.

There is minimal training involved in the use of the tablet. Basically, subjects were shown how to point and select a menu item from the tablet overlay; how to hold the stylus in an optimal position that would avoid accidentally pressing the button on the stylus. Depressing the button will generate arbitrary AutoCAD commands since it was not programmed to perform any particular function.

Recognition and selection accuracy

The recognition performance of the speech recogniser is based on the percentage of words correctly recognised on first-time verbalization. The performance during device template training was between 95-97%. For the test vocabulary training, mean recognition performance was 90.3%. This was based on single pass training as suggested in the manual. The speech usage performance was obtained by analysing the verbal content of speech during task performance using the same criterion, that is, first-time recognitions.

Speech performance, as discussed earlier, tends to drop during use and for this class of recogniser it ranges between 50-80% (see Section 7.2.1). The mean usage performance obtained was 79.7%. This supports the claims made by the manufacturer concerning PRONOUNCE; and this value (ie. 80%) was taken as the upper recognition threshold, that is, the optimal level for speech performance. Given that the vocabulary size for the CAD task would be increased by 4-5 times in subsequent experiments, a lower limit was set for usage performance, at 60% as opposed to 50% as said in the literature. This is because:

(1) the vocabulary size to be used will still be smaller (ie. <150 words) than other applications (eg. text processing) on which most studies are based;

(2) the vocabulary will comprise mainly single words, not phrases consisting of more than two words; and more important,

(3) to reduce the range in speech performance variability so that users may be encouraged, on the basis of frequent feedback, to try achieve the upper limit.

The claimed pointing accuracy of the tablet is 100%. This is based on a single pass pointing, that is, depressing the stylus once to select the menu item. This accuracy was confirmed in the test session. However, it should be pointed out that this test was carried out independent of a real task. Because of the sensitivity of the stylus tip, there is a general tendency for the pointing accuracy to be reduced in actual task performance.

Response speed

Response speed refers to the duration from stimulus onset to response offset. That is, the time taken to generate a user response, measured from the time a command card was displayed to the subject to the time the same command entry was completed by the subject. For example, if a card bearing the command "Hatch" was displayed, the subject was required to respond immediately by either verbalizing the same command once or pointing to it from an array of commands in the tablet menu. Each command in the list was shown twice in a random order.

The average response speed was 0.73 seconds and 2.04 seconds for speech and tablet input, respectively. This means that the response made via speech is quicker than by hand. A probable explanation for this is that manual input involves visual search of the tablet menu which delays the response; whilst speech input in this case simply involves word-repetition. This finding supports the general claim that speech input is relatively faster than manual input (see Chapter 2).

Ease of use

The speech recogniser was claimed to be easy to use by O2S1 but not by O2S2 due to the problems of verbal repeats of commands and voice fatigue with extended use. The graphics tablet, on the other hand, was claimed to be easy to use by both subjects though there were two general concerns. The first concerned the sensitivity of the stylus tip which led to many unwanted inputs, and the second relates to the organisation of the commands in the menu overlay which required considerable visual search.

Flexibility

With the speech recogniser, the cable of the headset microphone tended to constrain head movements and there was occasional slippage of the headset from its position during use. With the tablet, the cable of the stylus tended to obstruct visual search. In terms of tablet height, both subjects preferred the tablet to be moderately tilted at an angle as opposed to the low or high-tilt positions.

7.4.2 Application task : measures of system performance

The findings in this section are based on the optimisation of the system to a CAD task.

Time on task

The time taken to complete the task differed between input devices. Using the speech recogniser to perform the task took 20.5 minutes, whilst using the graphics tablet it took 15 minutes. This means that there is a tendency for the devices to differ in terms of overall time on task.

Product quality

Both input devices produced, on average, the same number of errors: 5.5 for the recogniser and 6.5 for the tablet. In other words, both input devices are comparable in terms of errors in the drawing.

Production cost (time)

The time taken to generate a drawing entity averaged 76.9 seconds using the recogniser, while using the tablet it averaged 56.3 seconds. This difference is considerable, suggesting that the devices may vary in terms of production cost (time).

Since no statistical analysis can be performed on the above results due to the small sample size, the variability in performance reflects individual differences. These differences, such as user satisfaction and preference, could be illustrated by data from the questionnaire.

7.4.3 Questionnaire findings

The role of the questionnaire (see Appendix 14) was to assess user acceptability of the system. Subjects were asked to rate their satisfaction with each input system and to indicate their preference, given a choice of the two devices.

Satisfaction ratings

The subjects differed in their satisfaction ratings of the recogniser. O2S1 rated it high (75%) while O2S2 rated it low (12%). With the tablet, both subjects had almost equivalent ratings: 61% and 63% for O2S1 and O2S2, respectively. For the speech input device, this finding illustrates a general problem relating to user attitudes and expectations, as discussed in Chapter 2.

Preference

Given a choice, O2S1 would prefer to use the recogniser to perform a CAD task because it is faster than the tablet. O2S2, on the other hand, preferred the tablet as it enables him to complete the task without the confusability problems he encountered in using the recogniser. Also, its use was less frustrating than the recogniser. This suggests that the ability to carry out the task with little interference from the device is crucial and this could determine user's preference for the system.

Assessment of strengths and limitations

Table 7.2 gives a summary of the strengths and limitations of each input device, based on the questionnaire.

TABLE 7.2

A summary of the strengths and limitations of input devices

Speech recogniser		**Graphics tablet**	
Strengths	*Limitations*	*Strengths*	*Limitations*
• quick	• voice fatigue with extended use • retraining of words due to confusability aspects of speech • recall of complex commands • frustrating due to recognition problems	• visible commands, less memory • ability to complete task without much problem	• familiarisation of command location in menu • difficulty in judging when the stylus is in contact with the tablet surface

On the whole, the questionnaires supported the observations made in the optimisation phases. In sum, it could be said that each device has problems of its own. These problems will be discussed in the next section.

7.5 DEVICE ASSESSMENT AND RECOMMENDATIONS

This section will present the main findings concerning each input device and the recommendations for modifying the demonstrator system.

7.5.1 Speech recogniser

Assessment

- The microphone, a SHURE-SM 10, was light but easily moved, which made it difficult to keep in any fixed position; and the wires leading from the microphone to the processor tended to constrain head movements.
- The headset was found to be uncomfortable, even for short periods of use (20-30 minutes). It should be realised that both users wore spectacles. Therefore, the combined use of a headset with spectacles may incur greater discomfort.
- The position of the microphone drastically affected the performance of the speech recogniser. A small movement away from the corner of the mouth caused a considerable fall in the activation of the device.

Recommendation

- **microphone**: optimum in its present form, except that the wires from the device may constitute a source of danger and may restrict considerable head movements.
- **headset**: this was not in its optimum form since it was uncomfortable. Adjustments to the grip should be made for each user.
- **microphone position**: this was optimum 1.5" at the corner of the mouth, NOT directly facing the mouth.

7.5.2 Graphics tablet

Assessment

- the tablet was found to be uncomfortable when placed flat on the desktop or when its default angle of tilt was at the maximum or minimum position.
- the stylus tip was too sensitive and tended to trigger off unwanted inputs; the wire leading from the stylus tended to obstruct user visibility, especially during selection and when the hand transits to the keyboard. The button on the stylus may be accidentally depressed causing more unnecessary inputs.
- the menu overlay was cluttered with symbols and commands that were not relevant to the task. Commands were not grouped for convenience; and the text size was too small for long commands (ie. >4 characters).

Recommendation

- **tablet height**: this was optimum at the mid-tilt position, approximately 15 inches from user's line of sight to the surface of the tablet.

• **stylus**: this was optimum in its present form, except that the cable from the device may obstruct hand movements, and the tip together with the button may generate unnecessary inputs if accidentally pressed.

• **menu overlay**: this was not optimum, particularly the size and style of text which slowed visual search. Bold text fonts should be used and the colour of the text should match those of the online speech vocabulary, that is, green. The organisation of the commands was not optimal. Commands should be grouped according to some logical principles that do not conflict with each other, such as functionality (eg. draw commands should be clustered together) and frequency of use (eg. line, circle, rectang, etc. should be contiguous to each other).

7.5.3 Other system aspects

Other modifications to the system that were deemed necessary include:

Design of speech vocabulary. This relates to the generation of two types of speech vocabulary: one online and another a hardcopy form (or offline). For both, the text font should match those used in the tablet menu in order to maintain consistency. Also, the organisation of commands in the hardcopy version should match those online. Given that the latter is constrained by the speech program, the arrangement follows an alphabetical order.

Design of screen information. The crosshair cursor was the same colour as the drawing elements, which hindered accurate selection of drawing entities. The crosshair should therefore be of a different colour from the basic drawing elements.

Design of menu program. This concerns the program for operating the tablet menu commands. Since some commands were not properly implemented, the program which was customised for this study should be checked and tested.

Design of task aids. This relates to the design of material to be used in the training and experimental sessions. It includes the design of instructions to subjects, user manual and CAD tasks. Since AutoCAD lacks consistency in the way some of its command rules are implemented, this should be made clear in the training instructions. For example, when the system responds with "objects or last or window" to commands such as **erase, copy, move**, etc., the general rule for "objects" is to select the object entity; the rule for "last" is to enter the command followed by **enter**; while the rule for "window" is to enter the command, define the window and press **enter**. Command ambiguity can slow learning, and hence performance.

7.5.4 Conclusion

On the basis of the findings concerning performance characteristics, the above suggestions were made to improve the use of the system for the subsequent experiments. Modifications of the demonstrator system concerned: (1) how different components of the system should be configured to ensure optimal performance; and (2) how task material, including tablet menu, speech vocabulary, instructions and screen information, should be designed to aid performance. The modifications serve as control measures on extraneous variables that can confound the main experimental effect.

7.6 SUMMARY

This study has highlighted two central issues: (1) the importance of optimising devices prior to use in order to establish their performance and to understand their characteristics; and (2) the need to understand user requirements of a system prior to use. The motivation for this optimisation is to ensure that the devices are representative of their class of input devices and that comparisons between classes are between equally good examples. This follows the optimisation of input devices by Long and Johnson (1982). Following the recommendations above, modifications of the demonstrator system were made accordingly.

NEXT CHAPTER HIGHLIGHTS

The demonstrator system derived from this optimisation study will be used in subsequent experimental investigations to evaluate the potential of speech input as a possible solution to the problems of unitary manual input. Chapter 8 describes an experiment aimed at identifying the problems of using unitary speech input. To this end, a comparative analysis of both input systems will be made based on behaviour and performance measures.

CHAPTER 8

Experiment 1: Assessment of Speech Input as a Unitary Solution to the Problems of Manual Input - Comparisons between Speech Recogniser and Graphics Tablet

Overview

8.1 Introduction

8.1.1 The solution - unitary speech input

8.1.2 Experimental aims and predictions

8.2 Method

8.2.1 Location and equipment

8.2.2 Subjects

8.2.3 CAD tasks

8.2.4 Experimental design

8.2.5 Procedure

8.3 Data analysis

8.3.1 Scoring of behaviour and performance

8.3.2 Blackboard models of system behaviour

8.4 Results

8.4.1 General

8.4.2 Effects of system on behaviour

8.4.3 Effects of system on performance

8.4.4 Correlation of behaviour and performance

8.4.5 Other questionnaire findings

8.4.6 Conclusion

8.5 Discussion

8.5.1 Comparisons of system behaviour models: unitary manual input versus unitary speech input systems

8.5.2 Assessment of strengths and limitations

8.5.3 Problems documented and possible solution

8.5.4 Design guidelines derived

8.5.5 Conclusion

8.6 Summary

Next chapter highlights

CHAPTER 8

Experiment 1: Assessment of Speech Input as a Unitary Solution to the Problems of Manual Input - Comparisons between Speech Recogniser and Graphics Tablet

OVERVIEW

In the observational study of CAD experts, it was shown that the use of unitary manual input devices resulted in non-optimal behaviours. This chapter investigates a solution to the problems of manual input by replacing it with unitary speech input. The main aim is to assess the suitability of unitary speech input based on analyses of behaviour and performance measures derived from naive users. The CAD system to be used in this experiment and subsequent investigations is derived from the optimisation study described in Chapter 7. Given that speech and tablet input have problems of their own, as stated in the previous chapter, the findings here show that the use of unitary speech input is also non-optimal.

8.1 INTRODUCTION

The problems of using unitary manual input, as documented in the study of CAD experts at work (Chapter 6), are that it resulted in considerable eye transitions off-screen and frequent hand transitions between input devices. These behaviours are considered here as non-optimal, as defined in Chapter 4. It is possible that both problems may be remediated by the use of a speech recognition device. The next section explains how and why unitary speech input might be a potential solution.

8.1.1 The solution - unitary speech input

There are two possible ways in which speech input could reduce non-optimal behaviours of manual input. Replacing the manual entry of commands and numerical data with spoken entry enables: (1) the eyes to gaze on the graphics screen for a greater proportion of the time on task, thus reducing the need for off-screen gazing, except in occasional circumstances such as checking the plan or CAD manual; and (2) the hand(s) to manipulate the graphics tablet for drawing operations only and the keyboard for entering textual data only, thus reducing the need to transit frequently between the tablet and the keyboard.

In the case of unitary speech input, the use of two input modes (ie. speech input=commands and numeric entry; manual input=graphical and textual data entry) offers certain advantages.

First, the distinct separation of input devices by modality (ie. recogniser=speech; tablet and keyboard=manual) which are compatible with each other, should help to ease learning. There is evidence (Wickens, et al., 1983; 1984) suggesting a unique compatibility relation between modalities of input (visual, auditory) and output (speech, manual), and codes of central processing (spatial versus verbal). According to Wickens et al. (1984), verbal tasks (eg. issuing commands) are best served by speech response and auditory inputs, whereas spatial tasks (eg. entering coordinates or drawing) are best served by visual-manual channels.

Second, the allocation of specific functions to input devices based on the function to which the device is best suited, should help to simplify the system, which in turn might ease learning. A system that is well-rationalised in terms of device functionality could reduce the problem of workload when making responses (Monk, 1986). The term *workload* includes memory load (ie. remembering device functions, etc.), manual load (ie. shifting of hands between input devices), and visual load (ie. attending to the device during use). Besides reducing the workload, a simplified system might benefit users, in particular naive users, leading to enhanced performance.

8.1.2 Experimental aims and predictions

The effect of unitary manual input on task performance, and the relationship between behaviour and performance, were not established in the observational study. Therefore, this experiment has two main goals:

(1) to investigate the suitability of unitary speech input as a solution to the problems of unitary manual input; and

(2) to document the nature and extent of the problems of speech input in CAD systems.

The above requires speech input to be compared with manual input. (Here, the term *manual input* will be used interchangeably with *tablet input.*) Therefore, there will be two systems involved: tablet input (or System A) and speech input (or System B). Tablet input will be the control condition and its predicted consequences for behaviour and performance will be based on previous empirical findings (Chapters 6 and 7). These are expressed below in IF-THEN statements, following the same format for knowledge description (see Figure 5.1, Chapter 5).

The disadvantages of using **System A** (Tablet input) over System B are:

IF

TASK is to input drawing commands and numerics
COMPUTER INPUT DEVICES are graphics tablet for entering commands, numerics and coordinates, and keyboard for entering text
OUTPUT DEVICES are graphics screen for displaying graphics and text screen for displaying text and system messages
USER is not able to select commands/numerics without directly looking at the tablet menu

THEN (behaviour)

- Moderate frequency and duration of eye gaze to the graphics screen.
- Moderate frequency and duration of eye gaze to the text screen.
- High frequency and duration of eye gaze to the graphics tablet.
- Moderate frequency and duration of eye gaze to the keyboard.
- Low frequency and duration of dominant hand being idle.
- Low frequency and duration of hand entering graphical data (drawing).
- High frequency and duration of hand inputting commands/data.

This would result in performance being:

- High product quality (ie. few drawing errors in the task output).
- High production costs (time) (ie. more time is needed to generate a drawing entity).
- Low production costs (efficiency) (ie. few commands required to produce an entity).
- Moderate user acceptability (ie. moderate ratings on performance and satisfaction by users).

The outcome is that behaviour and performance will be sub-optimal.

The advantages of using **System B** (Speech input) relative to System A are:

IF

TASK (as above)

COMPUTER INPUT DEVICES are speech recogniser for entering commands and numerics; graphics tablet for entering coordinates, and keyboard for entering text

OUTPUT DEVICES (as above)

USER is able to speak commands/numerics without the need to check for their recognition

THEN (Behaviour)

- Increase in frequency and duration of eye gaze to the graphics screen.
- Decrease in frequency and duration of eye gaze to the text screen.
- Decrease in frequency and duration of eye gaze to the graphics tablet.
- Decrease in frequency and duration of eye gaze to the keyboard.
- Increase in frequency and duration of dominant hand being idle.
- Increase in frequency and duration of hand entering graphical data.
- Decrease in frequency and duration of hand inputting commands.

This would result in performance being:

- High product quality.
- Low production costs (time).
- Moderate production costs (efficiency).
- High user acceptability.

The outcome is that behaviour and performance will be significantly improved over the manual input.

With the above goals and assumptions, the next section describes the experiment.

8.2 METHOD

8.2.1 Location and equipment

The experiment was conducted at the Ergonomics Unit, University College London. Unlike the optimisation study (Chapter 7), the setting for this and subsequent experiments was a normal university office room. Thus, there is potential for background noises from the external environment, such as mechanical drilling, voices, telephones ringing, etc. to influence the situation. Given this possibility, the voice setting for the speech recogniser was set at a moderate noise level.

The CAD system to be used in this experiment was described in Chapter 7; it is based on the optimised version of the demonstrator system. The general experimental set-up is shown in Figure 8.1. Figure 8.2 shows a subject using a speech recogniser to perform the task. The recording equipment (see Appendix 15) comprised: (1) a U-matic video cassette recorder; (2) two colour video cameras and tripods; (3) a digital timer and stopwatch; (4) a colour vision mixer; and (5) a video monitor to display the on-going recording.

8.2.2 Subjects

Due to the nature of the application domain, it was decided to use a single class of users, that is, inexperienced on CAD and/or the test software. But it was also necessary for the subjects to have sufficient motivation to be trained on the CAD system. Twenty-four volunteers served as subjects (14 male and 10 female) aged between 19 and 42 (mean=26) years, mostly university students (83%). All subjects had a minimum academic achievement of A-level. Fifteen of the subjects were British and nine non-British. Four subjects were left-handed, the rest were right-handed. Half of the sample had aided vision; they either wore glasses or contact lenses. Four subjects had minimal experience with other CAD systems but no subjects had used the test software. (Each subject was given an identity number, eg. E1S12 means Experiment 1 Subject 12.)

8.2.3 CAD tasks

The draughting task involved completing an unfinished office plan. This means the general office layout was given, and subjects were required to modify existing objects as well as create new ones following the hardcopy plans provided. There were three plans, showing detailed arrangements of office furniture. To enable generalisation of the findings, the selected plans were based on examples of real CAD work. The first plan (called Plan A), derived from AutoCAD's drawing library, was used in the training session (see Appendix 16). The remaining plans (Plan B and Plan C), were derived from the University's drawing office and were used in the experimental sessions (see Appendix 17).

Plans B and C showed the layout of the offices on the second floor of the Bedford Way Building, University College London, with details of the office furniture created for the purpose of this experiment. To control for task complexity, both plans were made to be as similar as possible in the type and number of drawing entities each contained.

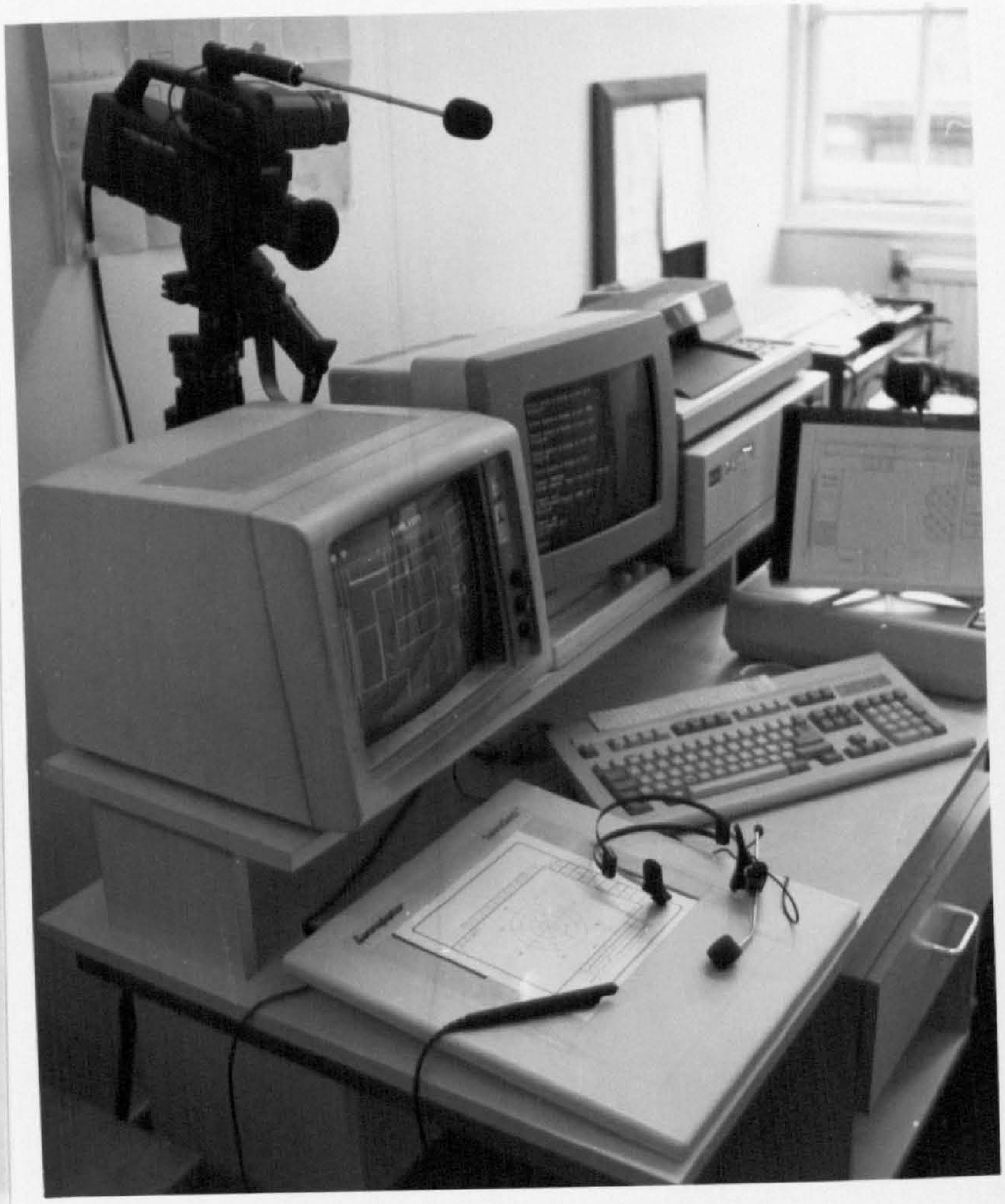


Figure 8.1. Configuration of Experimental System

Figure 8.2. General Experimental Set up showing a Subject using a Speech Recogniser to perform a CAD Task

8.2.4 Experimental design

The experiment employed a mixed 2 x 2 ANOVA design with one between-subject variable (the two input devices) and one within-subject variable (the two task plans). Half of the subjects were assigned to the Speech input group, and the other half to the Tablet input group, based on their performance scores obtained in the training session. This was a composite score of three measures: the number of drawing entities, the number of errors and task completion time. The usual precautions of counterbalancing the order of presentation of the tasks and allowing practice on the system were taken in the experimental session.

Behaviour and performance measures

The experiment yielded the following quantitative measures:

(1) behaviour measures: frequency and duration of eye gaze to I/O devices, plan and manual; frequency and duration of hand manipulation of graphics tablet and keyboard for data entry; and frequency of first-time recognitions, substitution and rejection errors as verbal content of speech; and

(2) performance measures: product quality, production costs and user acceptability.

These metrics were defined in Chapter 4. To recapitulate, *frequency* refers to the number of behaviour type per second; *duration* is the percentage of total time for each behaviour type, ie. relative duration.

Of the two measures, duration is considered more important towards understanding improvements in behaviour. This is because duration reflects the overall time spent for each behaviour type, whilst frequency denotes transitions or 'interruptions' in behaviour. It is therefore crucial to decrease the frequency of particular behaviour occurrence in order to increase the duration. This issue must be borne in mind in interpreting the results.

8.2.5 Procedure

The experiment was conducted in two sessions. The first for the familiarisation and training on AutoCAD; the second for the training on the test system and for performing the two experimental trials. The sessions were separated by at least one day; the maximum separation was 2 days. Each session lasted approximately 2.5 hours. The remainder of this section elaborates on the conduct of each session.

Training session

This session was conducted in three phases:

Phase 1. *Introduction to the experiment and CAD system* (15 minutes). Subjects were given an overview of the experiment and the instructions (see Appendix 18a). Then, they were introduced to the CAD system, followed by the completion of a profile form (Appendix 18b).

The form was used to obtain background information regarding the subjects and their motivation for doing the experiment.

Phase 2. *Learn AutoCAD using the keyboard* (60 minutes). In this phase, subjects were shown how AutoCAD works, using the keyboard for command entry and the tablet for coordinate entry. This is to eliminate the confounding effect of device-bias in the experimental session. Subjects were encouraged to ask questions during the demonstration.

Phase 3. *Practise doing a CAD task* (Plan A) *until completed*. This practice task was performed using the same input devices in Phase 2. Subjects were allowed to ask whenever in doubt or to refer to the manual whenever necessary. The time to complete the task was recorded with a stopwatch. At the end of this session, the content of the task output was assessed for quality and quantity of the drawing entities.

<u>Experimental session</u>

Each subject completed the following three phases.

Phase 1. *Train on input device and practise using the assigned device* (45 minutes). Subjects were first trained on how to use the input device. For the Speech input group, this includes template, user and device training (see Chapter 7). For the Tablet input group, it involves pointing to the menu items and remembering their locations in the menu (see Appendix 19a). Next, the subjects used the device to practise drawing for about 15 minutes. Problems in speech recognition were resolved for each subject, such as retraining the word, adjusting the microphone, etc.

Phase 2. *Perform two draughting tasks* (80 minutes). Each subject performed two tasks, separated by a 5-minute rest interval. The time allowed for each task was set at 40 minutes. Subject's behaviour was recorded on video for a duration of 15 minutes per task. Each recording commenced after subjects had performed about 10-12 minutes of the task. This segment of the task was chosen for recording to avoid any effects due to settling down to the task at the beginning, and any possible fatigue effects towards the end, particularly with Speech subjects.

Subjects were instructed to work accurately, but also as quickly as possible and were reminded that at some point they would be requested to stop even if they had not completed drawing. Speech input subjects were instructed to retrain a particular word that was not recognised sometimes and to remain as consistent as possible in their speech production. The recording form used in this session is given in Appendix 19b.

Phase 3. *Complete questionnaire* (20 minutes). At the end of the above session, subjects were debriefed concerning their experiences and were asked to complete a questionnaire (see Appendix 20). The purpose of this questionnaire was to obtain the following information: (1) subjective ratings on task performance, user satisfaction and device use; (2) problems, strengths and limitations of device use; and (3) the type of commands frequently-used and those considered difficult to use. This information is crucial in verifying the behaviour protocol and performance measures.

8.3 DATA ANALYSIS

8.3.1 Scoring of behaviour and performance

A continuous 14-minute segment of each videotape was scored. This was selected after one minute of the tape had elapsed.* The types of visual and manual behaviours scored are

* There are two issues concerning this data selection which are applicable across experiments. First, the same sample was taken from all the protocols, and thus has the primary virtue of consistency. It is possible, however, that this sample is not representative of the tape as a whole. Although there is no formal evidence on this point, informal analysis suggests that the behaviours scored are representative of those in the remainder of the tape that were not scored. Second, the duration of the tape to be scored may vary between experiments providing the selected segment is sufficient to enable a characterisation of the various behaviour types. This too can be gauged from an informal analysis of the tape and the results of the statistical analysis.

rential statistics (ANOVA, Pearson product-moment correlation) were performed on the data. The predictions in Section 8.1.2 were tested at the $p=0.05$ level. (Note: for significant results, the actual probability level will be reported.)

8.3.2 Blackboard models of system behaviour

The behaviour data were used to construct a model of system behaviour for each input type. The framework for the model was described in Chapter 5. This group model illustrates the behavioural knowledge used by naive users in carrying out the tasks. The models will be used in understanding why and how performance differed between systems.

8.4 RESULTS

8.4.1 General

The analysis in this section concerns: the homogeneity of the sample prior to group assignment; the effects of task plan; the effects of learning; and speech recognition performance.

Group equivalence

In order to test for the null hypothesis of no difference between the two device groupings as a result of assignment, a oneway ANOVA was performed on the performance scores obtained in the training session. The analysis revealed $F(1,22)=1.07$, $p>0.05$ to be not significant. This confirms that the sample was homogeneous prior to their assignment.

TABLE 8.1

Categories of Visual and Manual Behaviours

Visual Behaviour

Direction of Looking
1 Text screen
2 Graphics screen
3 Tablet menu
4 Tablet/Stylus
5 Keyboard
6 Drawing plan
7 CAD Manual
8 Speech list (online)
9 Speech list (hardcopy)
10 Experimenter
11 Elsewhere

Manual Behaviour

Which Hand
1 Right
2 Left
3 Both

Destination of Movement
1 Tablet
2 Stylus
3 Keyboard
4 Drawing plan
5 CAD Manual
6 Speech list (hardcopy)
7 Headset microphone
8 Screens
9 Writing object
10 Drinking mug
11 Elsewhere

What Movement
1 No movement/Idle
2 Locating menu item
3 Entering command
4 Inserting graphical data
5 Dragging crosshair
6 Pressing function key
7 Typing-in text/data
8 Pick up (object)
9 Put down (object)
10 Holding (object)
11 Pointing to screen/other

TABLE 8.2

Categories of Verbal Behaviours

Verbal Behaviour

Utterance type
1 Command/numeric data
2 Retrain word
3 Request help (how to do it, etc.)
4 Respond to experimenter's query
5 Self-commentary (task)
6 Self-commentary (non-task)

Command/Data entry
1 Single pass recognition
2 Repeat: Substitution error
3 Repeat: Rejection/Insertion error
4 Repeat: Forgetting error
5 Repeat: Commission error
6 Repeat: Inefficiency
7 Repeat: Incorrect input

Command/Data list
1 DISPLAY Zoom/Zoom all
2 DISPLAY Redraw
3 DISPLAY Zoom window
4 DISPLAY Zoom extent
5 DISPLAY Zoom previous
6 DRAW Arc
7 DRAW Circle
8 DRAW Fill
9 DRAW Line
10 DRAW Text
11 DRAW Trace
12 EDIT Break
13 EDIT Chamfer
14 EDIT Change
15 EDIT Copy
16 EDIT Erase
17 EDIT Fillet
18 EDIT Move
19 DISPLAY Pan
20 MODE Grid
21 MODE Ortho
22 MODE Osnap
23 MODE Snap
24 UTILITY Help
25 DOS Cancel
26 DOS Enter
27 DEVICE Goodbye
28 DEVICE Listen
29 DEVICE Pronounce
30 SubCommand aspect
31 SubCommand close
32 SubCommand distance
33 SubCommand drag
34 SubCommand ellipse
35 SubCommand last
36 SubCommand off
37 SubCommand on
38 SubCommand radius
39 SubCommand rectangle
40 SubCommand undo
41 SubCommand window
42 NUMERALS
43 UNITS feet/inches
44 SubCommand angle
45 Other

Effect of task plan

A oneway ANOVA carried out on the production cost (time) and product quality (error) data showed no significant effect of plan with results for production cost, being $F(1,46)=0.98$, $p>.05$; and product quality, $F(1,46)=0.56$, $p>.05$. Pooling both data sets also produced no effect. This confirms that both design problems were comparable in terms of the number of entities in the drawing and the level of complexity of the draughting task. But this was not perceived as such by subjects as revealed from the questionnaire (see Question 5, Appendix 20). 74% of the subjects tended to find one task to be more difficult than the other, while 26% estimated both tasks to be equally difficult. This indicates that perceived difficulty may not necessarily match actual difficulty.

Effect of learning

Generally, subjects performed better on the second task, whichever the design plan. A oneway ANOVA revealed a highly significant learning effect in terms of the percentage of entities that were completed, $F(1,46)=20.33$, $p<.01$. The error data just reached significance at the 5% alpha level, with $F=3.83$ and $df=1,46$. The learning effect was also extremely significant on the combined data - entity and error. It confirmed further that learning did occur over time. This, however, is not surprising given that subjects were inexperienced and the devices are novel to them. On the basis of this finding, all subsequent analyses used the second task data, unless it was necessary to examine both results together.

Speech recogniser performance

Speech recognition performance obtained during template training ranged between 82% and 90%. This was based on a single pass training. Only one subject (E1S5) attained a recognition rate of less than 60% (58%). The same subject continued to experience recognition difficulties during the experimental sessions. As defined in Chapter 7, recognition performance during use should range between 60% and 80%. Mean speech usage performance obtained for the first and second task was 65.7% and 63.5%, respectively. This small difference suggests that speech recognition tended to be the same over time. The severest confusion was between "Line" (a command) and "nine" (a number); and it affected most of the subjects.

8.4.2 Effects of system on behaviour

The purpose of this analysis is to test the predictions in Section 8.1.2 concerning the effects of system on behaviour. Only behaviour types that are considered relevant to task performance will be presented. A complete summary of the ANOVA results is given in Appendix 21.

Eye gaze to specific targets

Within a scored duration of 840 seconds, Speech subjects made, on average, 381 eye transitions (ie. one transition per 2.26 secs.), while Tablet subjects generated 453 eye transitions (ie. one transition per 1.90 secs.). This difference in eye gaze between groups was significant, $F(1,22)=5.67$, $p<.03$. Thus, Tablet subjects made more eye transitions than Speech subjects.

Table 8.3 summarises the group results (n=24) for a few, select visual behaviour types. To maintain coherence throughout, the duration results for each behaviour type will be presented first, followed by the frequency results.

It is evident from Table 8.3 that Tablet subjects spent longer looking at the graphics screen (52.5%) than Speech subjects (40.2%). Also, Tablet subjects gazed more frequently to the graphics screen (.21) than Speech subjects (.18). Speech subjects, on the other hand, looked longer at the text screen (44.1%) than Tablet subjects (14.9%). The frequency of gaze to this screen is .17 times per second for Speech subjects and .12 for Tablet subjects. Separate ANOVA tests performed on these data produced highly significant differences (see Table 8.3).

These results indicate that using unitary speech input leads to: (1) less time being spent looking at the graphics screen and less frequent eye transitions to this screen; (2) more time is spent gazing at the text screen and more frequent transitions too. This behaviour is therefore non-optimal and did not support the predictions in Section 8.1.2.

The difference in gaze between the two groups is also highly significant for the input devices. But it should be pointed out that the Speech group did not use the tablet for entering commands while the Tablet group did not use the keyboard for reentering commands. The keyboard is used by both groups for entering text. In light of this, the significant results must be interpreted with great caution.

As expected, the Tablet subjects spent more time looking at the graphics tablet to enter commands and numerics (21.3%). The frequency of eyes transiting to the tablet by this group is .14 per second. The differences in gaze are highly significant when analysed with ANOVA (see Table 8.3). Speech subjects, however, spent 2.7% of the time looking at the keyboard whilst Tablet subjects only 1.1%. This difference was not significant. The frequency in gaze between groups is too low to be meaningful, although the difference in the figures is highly significant when analysed with ANOVA (Table 8.3).

The above results suggest that using speech input leads to: (1) null eye gaze to the tablet; and (2) some eye gaze to the keyboard for reinputting commands. The findings tended to support the predictions in Section 8.1.2, that behaviour may be improved.

TABLE 8.3

Effects of Manual and Speech Input Systems onVisual Behaviour - Eye Gaze to Specific Targets (n=24)

	System A	System B	ANOVA results	
	mean frequency	*mean frequency*	*F(1,22)*	*p*
Eyes-gaze-Graphics screen	.21	.18	4.56	.04
Eyes-gaze-Text screen	.12	.17	8.48	.008
Eyes-gaze-Graphics tablet	.14	.0	261.15	.000
Eyes-gaze-Keyboard	.0	.03	33.22	.000
Eyes-gaze-Drawing plan	.06	.06	.31	.58
	% duration	*% duration*		
Eyes-gaze-Graphics screen	52.5	40.2	17.51	.0004
Eyes-gaze-Text screen	14.9	44.1	139.63	.000
Eyes-gaze-Graphics tablet	21.3	0.2	249.56	.000
Eyes-gaze-Keyboard	1.1	2.7	3.33	.08
Eyes-gaze-Drawing plan	9.3	8.4	.63	.44

System A=Tablet input; System B=Speech input

Looking at the plan as work progresses appears to be the same between the two groups, whether in terms of duration or frequency (Table 8.3). This indicates that the use of the drawing plan to aid performance is the same between subjects, independent of the device in use.

Hand manipulation of input devices

This analysis is to test the prediction that using speech input reduces the time spent operating the input devices as well as the frequency of use. The results will be based on the dominant hand used to manipulate the tablet. Table 8.4 summarises the group results (n=24) on manual behaviour types.

The duration of the hand being idle for Speech subjects is 65.9% of the time; with Tablet subjects it is only 39.7%. But the frequency of the hand being idle is .16 for Tablet subjects and .08 for Speech subjects. These differences in hand idleness between subjects were highly significant (see Table 8.4), suggesting that using speech input keeps the hand(s) less busy. This supports the prediction in Section 8.1.2 concerning hand use.

Looking at individual hand activity, Speech subjects spent 29.3% of the time entering graphical data (or drawing) and the frequency of this activity is .08 per second. Tablet subjects spent 41.8% of the time drawing and the frequency of this occurring is .10 per second. These differences were also found to be very significant (Table 8.4), indicating that using speech input reduces the duration and frequency of drawing. This finding did not support the predictions.

In terms of data entry (command and numeric), Speech subjects spent 1.6% of the time entering data via the keyboard. The frequency of this occurrence is .02 per second. Tablet subjects, on the other hand, spent longer periods of time entering data via the tablet and locating menu items (17.3%). The frequency of manipulating the tablet for this is .17 per second. The differences in these activities between groups are significant as shown in Table 8.4, indicating that the unitary use of speech input has resulted in significant keyboard use.

The findings here suggest that using speech input results in the hand: (1) being less busy overall; (2) spending less time drawing; and (3) spending more time entering data using the keyboard. The latter necessitates shifting the hand between the tablet and the keyboard. This behaviour is therefore considered non-optimal.

Verbal content of speech

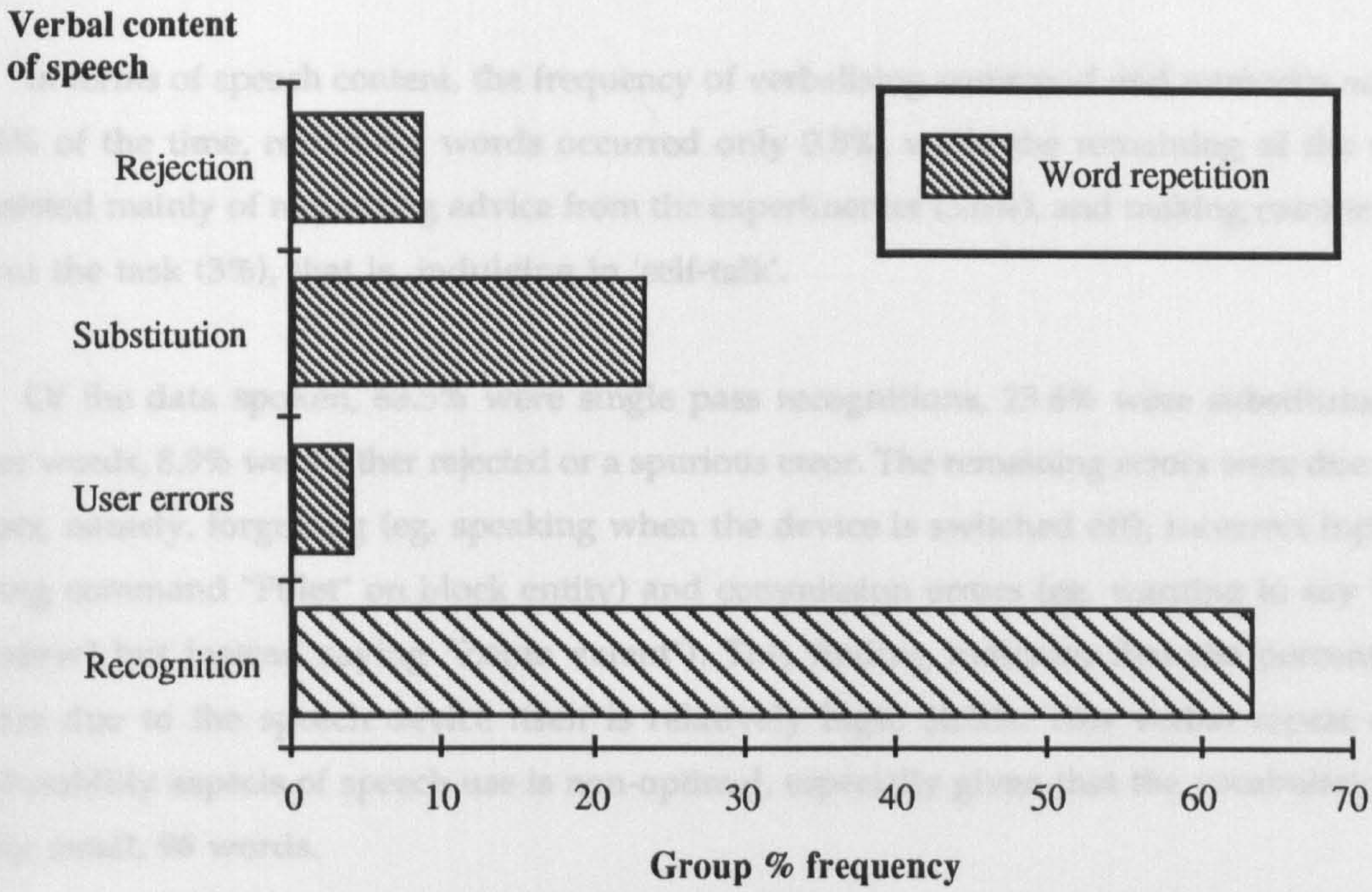
This analysis is to assess the nature of speech input use, with respect to the type of utterance made and the type of speech error incurred. Figure 8.3 summarises the verbal content of speech

TABLE 8.4

Effects of Unitary Manual and Speech Input Systems on Manual Behaviour - Patterns of HandUse (n=24)

	System A	System B	ANOVA results	
	mean frequency	*mean frequency*	*F(1,22)*	*p*
Hand-Idling	.16	.08	69.22	.000
Hand-Drawing	.10	.08	6.14	.02
Hand-Entering data	.0	.02	12.28	.002
Hand-Locating menu	.17	.0	458.06	.000
	% duration	*% duration*		
Hand-Idling	39.75	65.89	62.85	.000
Hand-Drawing	41.77	29.30	11.13	.003
Hand-Entering data	0.23	1.57	8.12	.009
Hand-Locating menu	17.30	.0	183.04	.000

System A=Tablet input; System B=Speech input



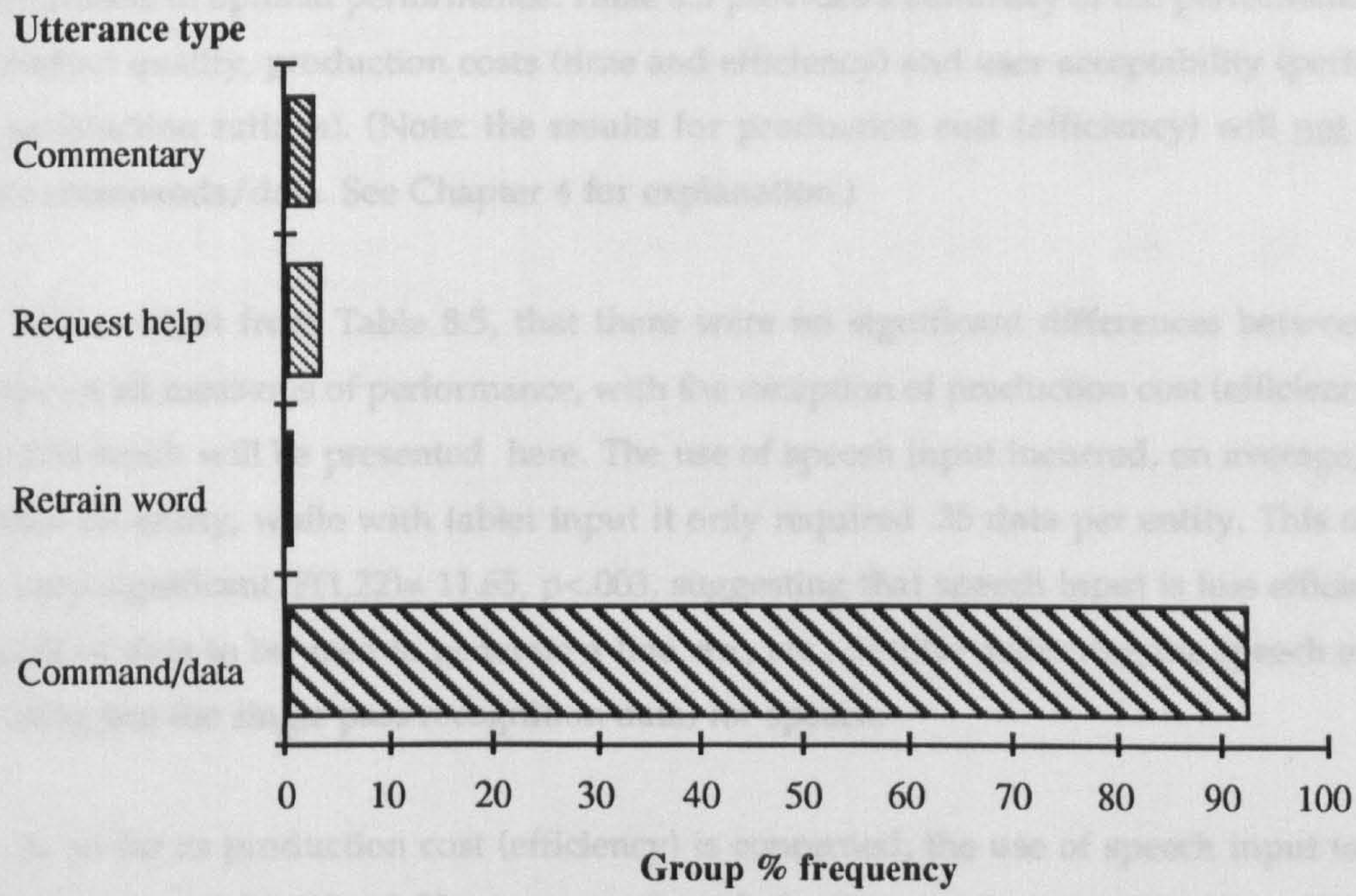


Figure 8.3. Speech Recognition Performance of Speech Input Group (n=12)

of the Speech input group. (The results are based on the second task, expressed as relative %.)

In terms of speech content, the frequency of verbalising command and numerics occurred 92.6% of the time, retraining words occurred only 0.8%, while the remaining of the speech consisted mainly of requesting advice from the experimenter (3.6%), and making commentaries about the task (3%), that is, indulging in 'self-talk'.

Of the data spoken, 63.5% were single pass recognitions, 23.6% were substituted with other words, 8.9% were either rejected or a spurious error. The remaining errors were due to user errors, namely, forgetting (eg. speaking when the device is switched off), incorrect input (eg. giving command "Fillet" on block entity) and commission errors (eg. wanting to say "Zoom window" but instead saying "Zoom extent"). This finding indicates that the percentage of errors due to the speech device itself is relatively high, 32.5%. This verbal repeat due to confusability aspects of speech use is non-optimal, especially given that the vocabulary size is fairly small, 96 words.

8.4.3 Effects of system on performance

The results in this section are aimed at assessing the extent to which the use of each input device results in optimal performance. Table 8.5 provides a summary of the performance results on product quality, production costs (time and efficiency) and user acceptability (performance and satisfaction ratings). (Note: the results for production cost (efficiency) will not indicate whole commands/data. See Chapter 4 for explanation.)

It is evident from Table 8.5, that there were no significant differences between device groups on all measures of performance, with the exception of production cost (efficiency). Thus, only this result will be presented here. The use of speech input incurred, on average, .45 data to draw an entity, while with tablet input it only required .35 data per entity. This difference was very significant, $F(1,22)= 11.65$, $p<.003$, suggesting that speech input is less efficient in the amount of data to be used to generate a line, arc, etc., despite deducting the speech error data (ie. using just the single pass recognition data) for speech.

In so far as production cost (efficiency) is concerned, the use of speech input to support performance is sub-optimal. However, on the whole, the use of manual input to support CAD performance is better.

8.4.4 Correlation of behaviour and performance

This analysis is intended to determine the relationship between behaviour and performance variables. Table 8.6 gives a summary of significant correlation results, for one-tailed proba-

TABLE 8.5

Effects of Unitary Manual and Speech Input Systems on Performance - Product Quality, Production Costs and User Acceptability

	System A	System B	ANOVA results	
	mean	*mean*	*F(1,22)*	*p*
Product quality	16.08	13.75	1.01	.33
Production cost (time)	8.28	8.11	1.51	.23
Production cost (efficiency)	.35	.45	11.65	.003
User acceptability (performance)	49.78	57.07	.84	.37
User acceptability (satisfaction)	59.52	67.26	.57	.46

TABLE 8.6

Correlation of Behaviour and Performance Measures: Experiment 1 results

System A : Tablet Input		
Behaviour with Performance variable	*r (12)*	*p*
Frequency of eye gaze to:		
All targets and production cost (time)	.52	.04
Graphics screen and production cost (time)	-.51	.05
All targets and production cost (efficiency)	-.75	.003
Graphics screen and production cost (efficiency)	.76	.002
Graphics tablet and production cost (efficiency)	.93	.000
Duration of eye gaze to:		
Graphics screen and production cost (time)	-.52	.04
Frequency of hand:		
Drawing and production cost (time)	-.65	.01
Entering data and production cost (time)	.52	.04
Idling and production cost (efficiency)	.92	.000
Locating menu and production cost (efficiency)	.82	.001
Duration of hand:		
Entering data and production cost (time)	.51	.05
Idling and production cost (efficiency)	.59	.02
Drawing and production cost (efficiency)	-.52	.04
System B: Speech Input		
Behaviour with Performance variable	*r (12)*	*p*
Frequency of eye gaze to:		
Keyboard and production cost (time)	.54	.04
Keyboard and production cost (efficiency)	.52	.04
Text screen and user acceptability (performance)	-.60	.02
All targets and user acceptability (satisfaction)	.54	.04
Duration of eye gaze to:		
Text screen and user acceptability (performance)	-.53	.04
Frequency of hand:		
Idling and production cost (time)	.49	.05
Drawing and production cost (efficiency)	.54	.03
Frequency of word:		
Recognition and production cost (efficiency)	.68	.007
Repetition and production cost (efficiency)	.75	.003
Substitution and production cost (efficiency)	.80	.001
Repetition and user acceptability (performance)	-.66	.01
Repetition and user acceptability (satisfaction)	-.52	.04
Substitution and user acceptability (performance)	-.50	.05

bility tests. Only a few, select results will be presented.

Speech input

The significant and positive correlations between eye gaze to keyboard with production costs (both time and efficiency) indicate that the frequent use of keyboard tended to increase the time required to generate a drawing entity and the amount of data needed to produce it (see Table 8.6). This implies two things: (1) the low level of skill in keyboard use; and (2) the disruption caused by having to shift to the keyboard. However, the more frequently the hand is involved in drawing, the more data are required to generate an entity (Table 8.6). This could mean that frequent drawing increases trial and error (as reported by subjects - see Section 8.4.5) in order to perfect the drawing, hence more data entry.

The correlations between frequency of single pass recognition and word repetition with production cost (efficiency) were significant and positive ($r(12)=.68$, $p=.007$ for recognition and $r(12)=.75$, $p=.003$ for repetition). This means that the more words get recognised or repeated, the more data are needed to draw, suggesting that frequent verbal repeat due to speech errors incurs cost, while increased verbal recognition encourages more data utilisation, also incurring cost.

The duration and frequency of eye gaze to the text screen correlates significantly and negatively with subjects' performance estimates ($r(12)=-.53$, $p<.04$ for duration and $r(12)=-.60$, $p<.02$ for frequency). This means that longer and more frequent eye gaze to this screen tends to compel subjects to rate their performance poorly, perhaps because of the tedious process of checking speech input recognition on the text screen. To support this: the correlations between frequency of word repeat with performance and satisfaction ratings were also negative, thus suggesting that increased word repetition did not satisfy the subjects, causing them to rate their performance low (see Table 8.6).

Manual input

The duration and frequency of gazing at the graphics screen correlates significantly and negatively with production time ($r(12)=-.52$, $p<.04$ for duration), and $r(12)=-.51$, $p<.05$ for frequency). This means that longer and increased eye gaze to this screen tended to reduce the time required per entity. In other words, this relationship is optimal. But frequent eye gaze to the same screen increased production efficiency costs (see Table 8.6). This indicates the more frequently eye gaze to (and, by implication, from) the graphics screen occurs, the more data are utilised to draw an entity. As in speech input, this could be due to trial and error on the subjects' part, as claimed by them (see Section 8.4.5).

The correlations between duration and frequency of data entry with production time were significant and positive (Table 8.6), indicating that high data input increases entity generation time. In other words, performance is slowed by increased data entry via the keyboard. On the other hand, frequent hand drawing activity reduces the time per entity drawn. Also, the longer period of time spent drawing tends to reduce the amount of data needed to produce a line (Table 8.6). Both the latter relationships are viewed as optimal in CAD performance. But the more frequent the hand is left idle, the more data are required in entity generation, thereby increasing production cost.

8.4.5 Other questionnaire findings

The role of this questionnaire (see Appendix 20) is to complement the behaviour protocols and to assess the role of the device in task performance. One subject (E1S13), however, did not complete the questionnaire satisfactorily, hence her data were excluded from this analysis.

Device usability ratings

ANOVA tests performed on the above ratings showed no significant differences in subjects' ratings of the device they used, in terms of its performance or in terms of ease of use and ease of learning. This means that both devices were well accepted, thus supporting the performance data in Section 8.4.3, and further confirms that the devices are good examples of their class of input devices as demonstrated in the optimisation study (Chapter 7).

Role of device in task performance

Despite the above ratings, 8.7% of the subjects (mainly Speech subjects) claimed that the device did not help to improve their task performance. The remaining (91.3%) found the devices contributed towards enhancing their performance. (Note: this result may be confounded by the fact that subjects compared the test device with using the keyboard during training.)

Speech subjects found the recogniser as aiding performance in the following ways:

- attention was less divided
- concentrate on screens and drawing
- more at ease and natural
- more enjoyment and a novelty
- repetition breeds familiarity with commands
- speeds up thinking process
- flexibility of device
- speaking is easier and faster than typing
- reduced time on task (no keyboard use)
- reduced keyboard use
- less tiring than typing

Tablet subjects, on the other hand, found the tablet to be:

- reduced manual reference because of menus on tablet

- no memory of commands
- grouping of commands eased use
- reduced time on task
- pointing easier than typing
- quicker than the keyboard
- reduced typing errors (eg. spelling)
- minimised keyboard use

However, there were problems in using these devices. For the Speech subjects, the problems relate to: (1) keeping their voices consistent, thus it became tiring to maintain the same tone of voice; (2) the confusability of some frequently-used commands and numbers, which led to a number of repeats; (3) background noise, particularly due to using the keyboard to reenter data; (4) frustration when instructions were misinterpreted; (5) discomfort of the headset microphone; (6) retraining words in the midst of task performance, thus disrupting task continuity; and (7) forgetting what to do caused by confusability.

The main problems experienced by Tablet subjects are: (1) positioning the stylus during entity selection; (2) stylus sensitivity and accidental pressing of the stylus button and/or tip; (3) familiarisation with the location of the menu items and remembering the meanings of the commands; (4) close proximity of commands in menu overlay which tends to cause a drift in item selection; (5) divided attention between tablet and text screen; (6) visual search of menu items; and (7) tablet size which tends to constrain hand movements.

AutoCAD learning and data configuration

Although the subjects had not used AutoCAD before, they found it easy to learn. There was no significant difference between subjects' ratings on AutoCAD learning, $F(1,21)= 0.50$, $p>.05$. Except for one subject (E1S3), all agreed that the training on AutoCAD was sufficient to enable them perform the tasks.

Two subjects (E1S3 and E1S6) found the commands were not easy to remember. Sixteen of the subjects could remember 75% of the commands (ie. about 72 commands), while 5 could remember 50% (ie. about 48 commands). Commands that are frequently used and those that subjects found difficult to use are summarised in Table 8.7. This breakdown of the data suggests two things.

First, the importance of this data type in CAD performance. Second, the contribution of this data type towards confusability. For example, in AutoCAD, the command "Enter" is frequently required to implement a data entry, but it is often confused with the subcommand "centre", thus increasing the frequency of its use. Similarly, the command "Line" is frequently needed to draw a line, but often it gets confused with the number "nine".

TABLE 8.7

Frequently-used and difficult commands (n=24)

20 Frequently-used Commands	*15 Difficult commands*
Line (23)	Chamfer (15)
Redraw (22)	Fillet (14)
Zoom (22)	Insert (10)
Copy (21)	Arc (9)
Move (21)	Osnap (9)
Erase (20)	Change (8)
Ortho (19)	Layer (8)
Arc (19)	Grid (7)
Cancel (17)	Trace (7)
Enter (16)	Text (6)
Trace (16)	Pan (6)
Text (13)	Snap (5)
Circle (12)	Move (5)
Grid (12)	Copy (4)
Insert (12)	Fill (4)
Break (11)	
Pan (10)	
Chamfer (8)	
Fillet (7)	
Snap (7)	

Note: Figures in brackets denote commonly-identified commands.

So, data configuration is crucial in the design of a speech vocabulary. Ideally, data that are easily confused either due to acoustically-similar characteristics or due to their very brief duration (eg. one, no) should be kept distinct and/or separate from each other. The findings here have shown that: (1) numbers tended to be confused with commands; and (2) frequently-used commands tended to be confused with less frequently-used ones. This issue will be discussed in Section 8.5.4.

8.4.6 Conclusion

The major hypothesis which predicts that the unitary use of speech input would enhance behaviour and performance is not supported in this experiment. On the contrary, its use has resulted in non-optimal behaviour and performance. The major behavioural problems were: (1) confusability aspects of speech input caused particularly by substitution errors; (2) between-screen eye transitions caused by the need to check for speech recognition on the text screen; and (3) between-device hand transitions caused by the need to reinput data using the keyboard. The performance problems relate particularly to production costs in terms of data inefficiency in entity generation. Despite these problems, speech input was well accepted by its users.

As predicted, manual input also resulted in sub-optimal behaviour and performance. The behavioural problems were mainly off-screen eye transitions to the tablet for entering data. Like speech input, the tablet was accepted as supporting task performance.

8.5 DISCUSSION

This section will summarise and discuss the central findings in four parts. The first part reviews the system models in terms of their content and operation. The second part compares the strengths and limitations of each input device. The third part discusses the problems documented with unitary speech input and identifies a possible solution to these problems. The last part summarises the design guidelines derived from this experiment to be used in the design of the next experiment.

8.5.1 Comparisons of system behaviour models: unitary manual input versus unitary speech input systems

The number of levels (STAM abstraction levels) in the models are the same, as explained in Chapter 5. In terms of the models' content, there are major differences between the systems. With speech input, there is greater recruitment of generative Task Specific KSs, in particular commands and numerics at the SubTask level of the blackboard. This is mainly due to the confusability aspects of speech input which required users to reinput more data than is necessary. In view of this, there is a tendency for users to use the keyboard as a backup device, which would mean recruiting knowledge at the Movement level. Also, there is a need to gaze more at the text screen to check for speech recognition feedback from the command interpreting process. Thus, compared with tablet input, there is greater recruitment of generative Tool Management KSs in the form of Text screen and Keyboard KSs.

With tablet input, there is more recruitment of graphical data at the SubTask level, which enabled users to spend more time drawing. This resulted in greater use of Graphics screen and Tablet KSs to perform the task. Because commands are entered via the tablet, there is a tendency to recruit more knowledge at the Movement level, for selecting the menu items from the tablet overlay. Hence, the use of generative Tool Management KSs, in the form of Tablet and Transducer KSs, is fairly substantial compared with speech input. Given that both systems use different input modes, with speech input there is a combined use of Eyes, Speech and Hand KSs; in the case of the tablet input, the knowledge consisted mainly of Eyes and Hand KSs.

The above implies that the allocation of resources to support performance occurs differently between systems. The allocation is managed by the knowledge executor which applies different control algorithms, taking into consideration the users' skills and history of KS use. With speech input, there is a tendency to rely more on the history of KS use, in terms of the frequency and duration of particular KS use (eg. Text screen KS). With tablet input, the user's skill in operating the tablet during menu selection appears to be an important factor. In particular, the ability to manipulate the device with or without visual monitoring.

The model, therefore, has helped to identify critical aspects of user behaviour using unitary speech or tablet input. Particular differences in knowledge recruitment are summarised in the next section.

8.5.2 Assessment of strengths and limitations

Behavioural assessment

In terms of the system behaviour model, the ***advantages*** of using System A (Manual input) vs. System B (Speech input) are:

(1) the recruitment of Eyes KSs to the graphics screen was for a longer duration and the frequency of this recruitment was high too. This enables the user to spend a significant proportion of the time on task conducting design activity. This behaviour is therefore optimal. But the increased frequency indicates frequent interruptions to graphics screen gazing caused by the need to gaze away at the graphics tablet for some aspects of the task, in particular command entry. This therefore may not be optimal to design behaviour.

(2) there was less frequent use of Eyes-Keyboard KSs, mainly for text entry. This therefore

led to less shifting of the hand(s) between the tablet and the keyboard.

(3) there was greater recruitment of Hand KSs for drawing purposes in terms of duration and frequency. Coupled with greater use of Eyes-Graphics screen KSs (see point 1), this led to an enhancement in design behaviour.

The *disadvantages* of using System A (vs. System B) are:

(1) the recruitment of Eyes-Graphics tablet KSs was exceptionally high, both in duration and frequency. This KS type was particularly needed for entering commands and numerical data from the tablet menu. Therefore, the time spent gazing off-screen and the frequency of such behaviour incurred substantial behavioural costs.

(2) the use of graphics tablet for data entry involved visual search of the menu items. This increased the use of Eyes-Tablet menu KSs for a significant proportion of the time, thus incurring costs to the user.

(3) there were longer periods of Hand KSs recruited to manipulate the input devices. This kept the hands generally busy during CAD performance, as commonly claimed in the literature (see Chapter 1).

Comparatively, System B (Speech input) has the following *advantages*:

(1) there was no recruitment of Eyes KSs to the graphics tablet because commands and numerics were spoken. In other words, there was no off-screen eye transitions to the tablet. This, then, resolved the problem of off-screen gazing incurred by manual input.

(2) the use of Hand KSs for manipulating the input devices was for shorter periods of the time on task. This means that the dominant hand was left idle for a significant proportion of the time, making CAD a less hands-busy task.

However, there are *disadvantages* of using System B:

(1) there was higher recruitment of Eyes KS to the text screen in terms of duration and frequency. The use thus incurred considerable time checking error messages and determining correct recognition of verbalised input. This in turn reduced the time spent on design activity itself.

(2) there was much use of Eyes and Hand KSs for reentering commands and numerics on the keyboard, in addition to text entry. This dependency on the keyboard as a backup device incurred production costs (efficiency). That is, more data were required to produce a drawing.

(3) the use of Hand KSs for drawing was for shorter periods of the time. Given that there was also less use of Eyes-Graphics screen KSs, this suggests that the time needed for drawing was taken up gazing at the text screen for command and recognition feedback. This behaviour is thus non-optimal.

Performance assessment

Both systems differed significantly in terms of production cost (efficiency). System A incurred less cost in the quantity of data necessary for producing the output compared with System B. In other words, an obvious disadvantage of using System B for performing CAD task is: it is not data efficient.

However, both systems are equal in performance with regard to: (1) the quality of the task output; (2) the time taken to produce a drawing (entity) output; and (3) user acceptability. Both systems were also found to be easy to use and to learn, which resulted in positive ratings on system performance. These subjective evaluations emphasised the strengths of each system and their potential in CAD use despite the many behavioural problems discussed above.

To summarise, the use of unitary speech input has helped to resolve the problem of off-screen gazing to the tablet incurred by unitary manual input. However, because of its own behavioural problems, its use did not resolve the problem of hand transition between the tablet and the keyboard. In other words, there are still discrepancies between the normative and observed system behaviours. In addition to the above discrepancy, the performative model shows that knowledge recruitment of Eyes-Graphics screen is still for shorter periods of the time on task. This is not optimal in terms of normative design behaviour. This non-optimal behaviour is caused by problems of unitary speech input to be discussed in the next section.

8.5.3 Problems documented and possible solution

This section will focus on the problems incurred by unitary speech input in terms of the device and user factors.

Confusability aspects of speech

(1) Device problem

With a reasonably large vocabulary consisting of commands and numerics, the tendency to confuse the inputs is greater because of the principle by which the recogniser operates, that is, acoustic-phonetic matching (see Chapter 2), and the length of allowable time per utterance (ie. between .5 to 2 seconds in duration). Both were recognised as problems by the subjects (see Appendix 22). Generally, commands were substituted with other acoustically-similar commands and numbers, and vice versa (see Section 8.4.5). Since most commands required the entry of parameters in the form of numerics, the rate of confusion was therefore proportional to the rate of command entry.

(2) User problem

The lack of experience with CAD led to some confusion on the users' part. In particular, their limited knowledge of how to use certain commands, and which commands to use for aspects of task led to verbal requests for help from the experimenter (see Figure 8.3). This verbal exchange in turn led to more speech confusion because of similar-sounding words in speech. The users' inexperience with the device (which led to it being switched on during non-use), and their reluctance to retrain mis- and non-recognised words (which led to further increase in verbal repeats) added to the confusability problems.

Eye and hand transitions

(1) Device problem

The confusability aspects of speech necessitated frequent use of the keyboard as backup facility for data reentry. This in turn created frequent eye and hand transitions to the keyboard, and unnecessary data input. In addition, it necessitated frequent checking of the text screen to ensure correct recognition and implementation of the data input. This led to increased between-screen eye transitions.

(2) User problem

The inability to remain consistent in speech productions and the lack of CAD experience necessitated frequent gazing of the text screen for system feedback, particularly speech recognition, selection and command feedback (see Chapter 3). Coupled with the problem of device confusability, subjects tended to check the text screen rather 'automatically'.

On the basis of these system behavioural problems, and because speech input is generally effective in supporting task performance, its integration with manual input could improve its overall utility in CAD. As revealed from this comparative investigation, the use of manual input has some advantages for behaviour and performance. Therefore, integrating it with speech input should optimise the functionality of each device in a single use. In other words, a possible solution to the problems of unitary speech and manual input is to integrate them within a single system. The benefits of this approach and the strategy involved will be explained in Chapter 9.

8.5.4 Design guidelines derived

The major findings concerning the use of both systems are applied to develop some human factors guidelines. These guidelines will be used in configuring speech-plus-manual input systems for subsequent experiments. The guidelines are recommendations about ways of integrating speech and manual input within a single CAD system. The guidelines took into consideration some of the problems experienced by both system users. As such, they may be applicable to certain user level - the novices and naive users of CAD systems. The guidelines are primarily intended to support the design of more flexible systems, which in turn would support user learning and reduce memory load. Chapter 11 describes in greater detail their development and validation.

The following guidelines were derived from this experiment.

Guideline 1 - *the vocabulary used for speech entry should be manageable by the user.*

In view of speech confusability, user memory and experience, it is crucial that the vocabulary for spoken data entry is predictable. A vocabulary is predictable when a user's choice of inputs at any time is small, so that the system will be more likely to make a correct match in interpreting an entry. It was established in this experiment that users were able to remember between 50-75 per cent of the vocabulary (ie. about 50-70 words).

Guideline 2 - *ensure that commands are kept separate from numeric data.*

Because of speech confusability, user inconsistency and time per utterance length, commands were generally found to be confused with sub-commands and numbers, and vice versa. Since numerics are fewer in number and tended to contain fewer characters than most CAD commands, their separation from commands might reduce confusability.

Guideline 3 - *ensure that the spoken entries for any transaction are phonetically distinct from one another.*

Words which are easily confused by the speech recogniser should be replaced with other dissimilar-sounding words. Also, users should be encouraged to substitute the confused words with familiar words (ie. natural to them) which do not sound similar.

Guideline 4 - *the functions of each input device should be well rationalised and clearly distinct from each other.*

Clearly defining the functions that each device supports helps to optimise the utility of each input device and to simplify its use. In this experiment, the keyboard was assigned for text entry, the tablet for graphical and numeric entry, and the speech recogniser for command and numeric entry. But because of speech confusability, the keyboard was also used for command and numeric reentry.

8.5.5 Conclusion

Comparisons between manual and speech input systems have shown that the unitary use of speech input has problems of its own and is not a solution to the problems of manual input, given its technological constraints. However, its potential as an accurate and cost-effective input device, and its general acceptance by naive users suggests it could be combined with manual input. The design of an integrated speech-manual CAD system will depend on factors such as the characteristics of the task (size and configuration of vocabulary, etc.) and the functionality of the input device.

8.6 SUMMARY

This investigation of unitary speech input as a solution to the problems of manual input has demonstrated that it is not viable due to the behavioural and performance costs that its use incurred. In particular, the problems relate to speech confusability, between- and off-screen eye transitions as well as between-device hand transitions. Because of its role in supporting task performance and its general acceptance, speech input has potential in CAD. The system behavioural problems documented here may be alleviated by integrating speech with manual input in a single system. This should optimise the utility and functionality of each input device given that the devices are complementary and are similar in performance.

NEXT CHAPTER HIGHLIGHTS

Chapter 9 will investigate integrating speech and manual input as a potential solution to the problems documented in Experiment 1. Comparisons between different integrated CAD systems and a unitary speech system will be made using the same behaviour and performance measures used here.

CHAPTER 9

Experiment 2: Assessment of Integrating Speech and Manual Input as a Potential Solution to the Problems of Unitary Speech Input - Comparisons of Device-Task Allocation Strategies

Overview

9.1 Introduction

- 9.1.1 The solution - integrating speech and manual input
- 9.1.2 The strategy - function allocation of device to task aspects: frequency of use principle
- 9.1.3 Experimental aims and predictions

9.2 Method

- 9.2.1 Location and equipment
- 9.2.2 Subjects
- 9.2.3 CAD tasks
- 9.2.4 Experimental design
- 9.2.5 Procedure

9.3 Data analysis

- 9.3.1 Scoring of behaviour and performance
- 9.3.2 Blackboard models of system behaviour

9.4 Results

- 9.4.1 General
- 9.4.2 Effects of system on behaviour
- 9.4.3 Effects of system on performance
- 9.4.4 Correlation of behaviour and performance
- 9.4.5 Other questionnaire findings
- 9.4.6 Conclusion

9.5 Discussion

- 9.5.1 Comparisons of system behaviour models: unitary speech versus integrated speech-manual input systems
- 9.5.2 Assessment of strengths and limitations
- 9.5.3 Problems documented and possible solutions
- 9.5.4 Design guidelines derived
- 9.5.5 Conclusion

9.6 Summary

Next chapter highlights

CHAPTER 9

Experiment 2: Assessment of Integrating Speech and Manual Input as a Potential Solution to the Problems of Unitary Speech Input - Comparisons of Device-Task Allocation Strategies

OVERVIEW

In the previous chapter, it was demonstrated that the unitary use of speech input resulted in non-optimal behaviour and performance. Given that both speech and manual inputs are accurate and cost-effective in supporting task performance, their integration as a single system might improve behaviour and lead to enhanced performance. This chapter reports an experiment to investigate the potential of integrating speech and manual input as a solution to the problems of unitary speech use. By allocating each device to aspects of the task, based on the frequency of use principle, it was shown that the problem of non-optimal behaviour is greatly reduced.

9.1 INTRODUCTION

Due largely to the confusability aspects of speech, the unitary use of speech input has resulted in considerable between-screen eye transitions and between-device hand movements (Chapter 8). Given the present technological constraints of commercially-available speech devices, the problem of non-optimal behaviour may be alleviated by integrating speech and manual input. This combination is indeed promising, enabling bimodal interaction. The next section discusses why this is a promising solution.

9.1.1 The solution - integrating speech and manual input

The benefits of integrating input devices that are complementary and compatible to each other were discussed in Chapter 8. Integration as a solution (as opposed to a unitary solution involving only one device) has its advantages particularly if the input devices are similar in performance. This is because combining them should not degrade task performance. Instead, an integrated system that is well rationalised in terms of how the different devices might be combined should help: (1) to simplify the system; and (2) to optimise the functionality of each input device within the system. Therefore, an integrated CAD system (used interchangeably here with the term hybrid system), is one that combines two input modes in carrying out tasks. In short, it is a bimodal system that enables communication via different user and computer input modes (see Chapter 4 for this characterisation).

9.1.2 The strategy - function allocation of device to task aspects: frequency of use principle

Function allocation has always been central in the design of any human-machine system (Greenstein & Lam, 1985). System designers distinguish between function and task. A *function* is defined as "a general means or action by which the system fulfills its requirements" (DeGreene, 1970, p. 21). A *task* is described at the behavioural level, and is construed to be a composite of discriminatory-decision-motor activities performed by an individual, and directed toward accomplishing a specific amount of work within a specific work context (DeGreene, 1970). Here, device functions include information input, selection, positioning and moving (Chapter 2), while aspects of tasks include information entry, such as commands, text, parameters, etc. (Chapter 3).

Function analysis and task allocation between human and machine is widely discussed in the literature (eg. Clegg et al., 1989; Kantowitz & Sorkin, 1987; Greenstein & Lam, 1985; Bailey, 1982; Chapanis, 1965). The general consensus is that an integrated system with computers and humans should take advantage of both sets of characteristics to achieve better operation (Clegg et al., 1989). In other words, the relationship between user and computer should be complementary. Applying this notion of complementarity to the CAD system here means devices should complement and are compatible with each other and with the user. To achieve this, devices should be assigned flexibly to perform the functions to which they are best suited. This involves identifying and analysing various device functions plus the features of CAD tasks, then allocating the device to specific aspects of the task. This approach will be driven by knowledge of human factors and technological constraints, derived from previous empirical analyses of device use.

There were four guidelines derived from Experiment 1 (Chapter 8). Guideline 1 states that the vocabulary should be manageable by the user. As a command-driven system, commands constitute a major input in CAD. The command language itself comprises major commands and subcommands. Other data types includes graphical, numeric (ie. parameters) and textual data. In short, the task vocabulary can be fairly large. Given the limited capacity of users, the vocabulary should be divided, allocating some to speech input and some to manual input. In Experiment 1, users could remember between 50-70 commands. This, then, could be used as a criterion in determining the size of a vocabulary. This strategy is in line with guidelines on dialogue (command) design (eg. Cole et al., 1987; Bailey, 1982).

Guidelines 2 and 3 are concerned with data configurability. That is, commands should be distinctly separate from numerics (Guideline 2) and the spoken entries should be phonetically distinct from one another (Guideline 3). Both guidelines will be dealt with together. In Experiment 1, it was found that commands were generally confused with numbers and subcommands, and vice versa. Because commands constitute a larger subset of the data entry than numerics, separating commands from the rest of the subsets should reduce the rate of word confusion. The criterion for dividing the data set is the frequency of use principle. This has been suggested as a logical principle for grouping items (McKenzie, 1988; Cole et al., 1987). From Experiment 1, it was also learnt that some commands are more frequently used than others. Therefore, the command set could be classified into *high* and *low frequency* commands and subcommands.

The problem of similar-sounding words will be resolved for each individual user during template training. The trained words will be checked and easily-confused words (eg. "Enter" and "centre") will be replaced with acoustically-dissimilar words (eg. "Return" for "Enter").

Guideline 4 involves identifying the functions for which each input device is best suited. This issue was also raised by Whitefield (1986a). It is crucial that the combination is not just of input devices that are complementary and compatible to each other, but the devices are suited to perform those functions assigned to them. The various device functions were described in Chapters 2 and 3. The studies conducted so far have confirmed three main functions of input devices: for information input, entity selection and positioning/moving. Therefore, assignment of devices to task functions could use this as a basis.

The above guidelines constitute what is termed here a *device-task allocation* strategy. This means allocating input devices to particular aspects of the task. As a result of this approach, two alternative configurations of an integrated speech-manual CAD system were derived. The unitary speech system will be the baseline condition for comparing the performance of the alternative integrated systems. This unitary system (**System C**) allocates speech input to all data types: (1) high frequency commands (HFC) and related subcommands; (2) low frequency commands (LFC) and related subcommands; (3) numerical data (ND); and (4) basic words for operating the speech device (BW). This means 100% of the total data set (with the exception of graphical and textual data) will be allocated to speech input alone. This condition is the same as the Speech input (System B) condition of Experiment 1.

The first alternative system (called **System D**) allocates HFC and BW (ie. 45% of the total data set) to speech input, and LFC and ND (ie. 55% of the data) to tablet input. The second system (called **System E**) allocates LFC, ND and BW (ie. 59% of the data set) to speech input while HFC (ie. 41% of the data) to tablet input. (Note: BW has to be assigned to speech input only.) In all three systems, the graphical data will be allocated to the tablet due to its appropriateness as a drawing tool, and the textual data to the keyboard given that it is the best method for text entry (see designers' comments in Chapter 6).

9.1.3 Experimental aims and predictions

This experiment is intended to address the following:

(1) to investigate the potential of integrating speech and manual input as a single system based on the proposed strategy, and comparing these alternative integrated systems with a unitary speech input system; and

(2) to document the nature of the problems in using the integrated systems.

As in Experiment 1, the predictions concerning the effect of systems on behaviour and performance will be expressed in condition-action links (excluding the components that are kept constant throughout the conditions - output devices, the input devices for graphical and text entry). User acceptability is excluded from this analysis because of a methodological flaw in the administration of the rating scale.

Taking **System C** (unitary speech input) as a control condition, and using the findings of Experiment 1, the disadvantages of using this system over Systems D and E (integrated speech-manual) are:

IF
TASK is to input information for conducting design activity
COMPUTER INPUT DEVICE is speech recogniser for entering HFC, LFC, ND and BW
USER is able to look at the screens while speaking

THEN (behaviour)

- Low frequency and duration of eye gaze to the graphics screen.
- High frequency and duration of eye gaze to the text screen.
- Low frequency and duration of eye gaze to the graphics tablet.
- Moderate frequency and duration of eye gaze to the keyboard.
- High frequency and duration of dominant hand being idle.
- Low frequency and duration of hand entering graphical data.
- Moderate frequency and duration of hand entering commands/data.
- High frequency of word repetitions relative to word recognitions.

This would result in performance being:

- Moderate product quality.
- High production costs (time and efficiency).

The outcome is that behaviour and performance will be sub-optimal.

The advantages of using **System D** (integrated speech-manual system 1) relative to System C (unitary speech input) are:

IF
TASK is to input information for conducting design activity
COMPUTER INPUT DEVICES are speech recogniser for entering HFC and BW; graphics tablet for entering LFC and ND; and
USER is able to look at the screens while speaking <u>but</u> is not able to enter tablet menu items without directly looking

THEN (behaviour)

- Increase in frequency and duration of eye gaze to the graphics screen.
- Decrease in frequency and duration of eye gaze to the text screen.
- Increase in frequency and duration of eye gaze to the graphics tablet.
- Decrease in frequency and duration of eye gaze to the keyboard.
- Decrease in frequency and duration of dominant hand being idle.
- Increase in frequency and duration of hand entering graphical data.
- Increase in frequency and duration of hand entering commands/data.
- Decrease in frequency of word repetitions relative to word recognitions.

This would result in performance being:

- High product quality.
- Low production costs (time and efficiency).

The outcome is that behaviour and performance will be significantly improved over System C.

The advantages of using **System E** (integrated speech-manual system 2) relative to System C are:

IF

TASK is to input information for conducting design activity
COMPUTER INPUT DEVICES are speech recogniser for entering LFC, ND and BW; graphics tablet for entering HFC; and
USER is able to look at the screens while speaking <u>but</u> is not able to enter tablet menu items without directly looking

THEN (behaviour)

- Increase in frequency and duration of eye gaze to the graphics screen.
- Decrease in frequency and duration of eye gaze to the text screen.
- Increase in frequency and duration of eye gaze to the graphics tablet.
- Decrease in frequency and duration of eye gaze to the keyboard.
- Decrease in frequency and duration of dominant hand being idle.
- Increase in frequency and duration of hand entering graphical data.
- Increase in frequency and duration of hand entering commands/data.
- Decrease in frequency of word repetitions relative to word recognitions.

This would result in performance being:

- High product quality.
- Low production costs (time and efficiency).

The outcome is that behaviour and performance will be significantly improved over System C.

9.2 METHOD

9.2.1 Location and equipment

This experiment employed the same setting, demonstrator system, experimental setup and recording equipment of Experiment 1 (Chapter 8). Changes however were made to the design of the menu overlays and speech lists - online and offline (ie. hardcopy) - to suit the experimental requirements.

9.2.2 Subjects

The sample comprised 24 volunteers, 13 male and 11 female aged between 18 and 45 years, with mean age of 26.5 years. Fourteen of the subjects were university students, 10 were employees from various work settings. All subjects had a minimum qualification of A-levels. Thirteen of the subjects were British, 11 were non-British. Only one subject was left-handed; 14 had normal vision while 10 used glasses or contact lenses.

With the exception of 2 subjects who had no computer experience, the rest had some (<3 months) or a lot (>3 months) of experience. Half of the sample had some CAD experience, while the remaining half had none at all. On this criterion alone, the sample could be classified into: 50% novices and 50% naive or first-time users of CAD. Eight of the novices had participated in Experiment 1 while 4 had used other CAD systems. The assignment of novices and naive users to the experimental conditions was evenly split. This was to eliminate the possible effect of bias in subject grouping. Each subject was assigned an identity number (eg. E2S21).

9.2.3 CAD tasks

Like Experiment 1, the tasks were based on real CAD tasks. They involved draughting kitchen layouts using the plans provided. The plans showed detailed arrangement of kitchen items, furniture, etc., based on models of kitchen from brochures of kitchen manufacturers. These were adapted for the purposes of this experiment. Three plans were used; the first plan was used in the training session (Appendix 23), the second and third plans (see Appendix 24) were employed in the experimental sessions. Both experimental plans had the same number of drawing entities (65) and objects (36). This means the use of different types of commands should be the same in both tasks (eg. 8 lines, 5 rectangles, 1 trace, 4 fillets, etc.).

There were 3 speech vocabulary lists (named after the systems): List C contained a total of 112 words (95 HFC plus LFC/subcommands and 17 numbers) for use with System C. List D had 51 words (all HFC/subcommands) and List E had 66 words (49 LFC/subcommands and 17 numbers) to be used with Systems D and E, respectively.

9.2.4 Experimental design

The design was a oneway mixed ANOVA design with all subjects tested in the baseline condition (within subjects), and one half in the first experimental condition, while the other half in the second experimental condition (between subjects). The independent variable was system type with three levels (two experimental and one baseline conditions). The dependent variables were behaviour and performance measures. Assignment of subjects to the experimental conditions was based on their performance scores obtained in the training session and their CAD experience. The order of task presentation was counterbalanced across conditions and task plans.

Behaviour and performance measures

The same metrics used in Experiment 1 were used in this experiment, namely:

(1) behaviour measures: frequency and duration of eye gaze to I/O devices, plan and speech list; frequency and duration of hand manipulation of graphics tablet and keyboard; and frequency of single pass recognitions and verbal repeats due to substitution and rejection errors as verbal content of speech; and

(2) performance measures: product quality and production costs.

9.2.5 Procedure

As in the previous experiment, this investigation was conducted in two sessions: a training and an experimental session. The sessions were separated by between 1 and 2 days. The training session lasted approximately 2.5 hours and the experimental session about 2 hours.

Training session

This session was conducted in four phases:

Phase 1. *Introduction to the experiment and CAD system* (15 minutes). The procedure for conducting this session was the same as in Experiment 1. Likewise, the introduction, instruction material and profile form used here were the same as those used previously. The only addition to the training material was the manual which contained examples of hatching patterns derived from AutoCAD's pattern library. Subjects were asked about their reasons for participating in order to ensure that they had sufficient motivation to be trained on the system.

Phase 2. *Learn AutoCAD using the tablet* (50 minutes). The subjects were trained in AutoCAD using the graphics tablet for all data entry, except text which was reserved for the keyboard. This is to minimise the possible effect of negative transfer of learning, as reported by some subjects in Experiment 1 (see Appendix 22).

Phase 3. *Practise doing a CAD task* (40 minutes). This practice task was performed using the tablet and keyboard as above. Subjects were allowed to ask whenever in doubt or to refer to the manual whenever necessary. The time to complete the task was set at 40 minutes. At the end of this session, the content of the task output was assessed for the number of errors and drawing entities it contained.

Phase 4. *Train on speech recogniser* (45 minutes). Subjects were first trained on how to use the recogniser (user and system training). This was followed by template training of two speech lists, one for the baseline condition (List C), the other for the experimental condition (List D/E). Only List C was checked for word recognition. Words with a score greater than 8 were retrained and the check procedure was repeated for these words.

Experimental session

Each subject completed the following three phases.

Phase 1. *Practise using the recogniser* (20 minutes). Subjects practised briefly using the speech list to be used in the first condition. Problems in recognition were resolved for each subject, such as retraining the word.

Phase 2. *Perform two draughting tasks* (70 minutes). Each subject performed two tasks, separated by a 5-minute rest interval. The time allowed for each task was set at 35 minutes. Subject's behaviour was recorded on video for a duration of 15 minutes per task. The same recording procedure and task instructions used in Experiment 1 were applied here.

Phase 3. *Complete questionnaire* (20 minutes). This session concluded with subjects completing a questionnaire (see Appendix 25). The role of this questionnaire was to elicit information relating to system use, problems experienced with each system, strategy for overcoming the problems, system preference and commands frequently used.

9.3 DATA ANALYSIS

9.3.1 Scoring of behaviour and performance

The recorded behaviour protocol provided the source for the behaviour data, while the drawings and questionnaire provided the performance data. Tables 9.1a and 9.1b summarise the behaviour types - manual and verbal - scored from each subject's protocol. (The scoring menu for visual behaviour is the same as in Experiment 1, Table 8.1.) Using VITAS continuous scoring method, an 8-minute segment was selected after 4 minutes of the tape had elapsed. Like Experiment 1, steps were taken to identify the behaviour types prior to scoring. This is to ensure that the scoring was reliable.

Scoring of each subject's drawing was done online, that is, each drawing was retrieved and displayed on-screen, then analysed in detail for the types of errors made and the number of drawing entities in the drawing. The questionnaire was coded and processed using SPSS/PC+.

The data met the requirements for parametric tests. Thus, oneway ANOVA was used to test the predictions (Section 9.1.3) and Pearson correlation to examine the relationship between behaviour and performance. To test if two groups differed on any of the measures, Scheffe´ test for multiple comparison was performed on the means data. This procedure is more conservative than other post-hoc tests, such as Newman-Keuls, Tukey, etc. (see SPSS/PC+ Manual, 1988), thus any differences between the main effects (ie. system types) are therefore real differences. All predictions were tested at the 5% alpha level but for significant results, the actual probability levels will be reported.

TABLE 9.1a.

Categories of Manual and Verbal Behaviours

Manual Behaviour

Which Hand
1 Right
2 Left
3 Both

Destination of Movement
1 Tablet
2 Stylus
3 Keyboard
4 Drawing Plan
5 Command list
6 Headset microphone
7 Graphics screen
8 Text screen
9 Body
10 Drinking mug
11 Elsewhere

What Movement
1 No movement/idle
2 Drawing entity
3 Locating menu
4 Entering command
5 Entering numeric
6 Pressing function keys
7 Typing-in text
8 Calculating
9 Holding/Touching
10 Browsing list/manual
11 Adjusting microphone

Verbal Behaviour

Utterance Type
1 Command (first-time)
2 Command (repeat)
3 SubCommand (first-time)
4 SubCommand (repeat)
5 Numeral (first time)
6 Numeral (repeat)
7 Basic word (first time)
8 Basic word (repeat)
9 Retrain word

Recognition Status
1 Recognised (single pass)
2 Not recognised (substitution)
3 Not recognised (rejection)
4 Not recognised (forgetting)
5 Not recognised (wrong input)

Command/Data list
* see Table 9.1b

TABLE 9.1b

Command/Data list for Verbal Behaviour

System C	System D	System E
Command/data list	Command/data list	Command/data list
1 HFC Grid	1 HFC Grid	1 LFC Axis
2 HFC Ortho	2 HFC Ortho	2 LFC Osnap
3 HFC Arc	3 HFC Arc	3 LFC Status
4 HFC Circle	4 HFC Circle	4 LFC Chamfer
5 HFC Fill	5 HFC Fill	5 LFC Oops
6 HFC Line	6 HFC Line	6 LFC Limits
7 HFC Text	7 HFC Text	7 LFC Change
8 HFC Move	8 HFC Trace	8 LFC Break
9 HFC Copy	9 HFC Move	9 LFC Hatch
10 HFC Erase	10 HFC Copy	10 LFC Point
11 HFC Pan	11 HFC Erase	11 LFC Layer
12 HFC Redraw	12 HFC Pan	12 LFC Save
13 HFC Fillet	13 HFC Redraw	13 LFC Quit
14 Enter/Return	14 HFC Regen	14 LFC Array
15 HFC Cancel	15 HFC Fillet	15 LFC Ltscale
16 HFC Snap	16 HFC Enter/Return	16 BW Listen
17 LFC Chamfer	17 HFC Cancel	17 BW Goodbye
18 LFC Oops	18 HFC Snap	18 BW Pronounce
19 LFC Change	19 BW Listen	19 SubCommand on
20 LFC Break	20 BW Goodbye	20 SubCommand off
21 LFC Hatch	21 BW Pronounce	21 SubCommand aspect
22 LFC Layer	22 SubCommand on	22 SubCommand centre
23 BW Listen	23 SubCommand off	23 SubCommand endpoint
24 BW Goodbye	24 SubCommand aspect	24 SubCommand midpoint
25 BW Pronounce	25 SubCommand Erase last	25 SubCommand near
26 SubCommand on	26 SubCommand Erase window	26 SubCommand perpendicular
27 SubCommand off	27 SubCommand endpoint	27 SubCommand distance
28 SubCommand endpoint	28 SubCommand centre	28 SubCommand window
29 SubCommand last	29 SubCommand diameter	29 SubCommand last
30 SubCommand window	30 SubCommand 2-point Circle	30 SubCommand firstpoint
31 SubCommand Erase last	31 SubCommand 3-point Circle	31 SubCommand plus
32 SubCommand 2-point Circle	32 SubCommand close	32 SubCommand minus
33 SubCommand close	33 SubCommand undo	33 SubCommand set
34 SubCommand undo	34 SubCommand Text aligned	34 SubCommand ?
35 SubCommand Text aligned	35 SubCommand Text centred	35 SubCommand yes
36 SubCommand ellipse	36 SubCommand style	36 SubCommand no
37 SubCommand rectang	37 SubCommand ellipse	37 SubCommand rectangular
38 SubCommand drag	38 SubCommand rectang	38 Number 0
39 SubCommand Zoom all	39 SubCommand drag	39 Number 1
40 SubCommand Zoom previous	40 SubCommand last	40 Number 2
41 SubCommand Zoom window	41 SubCommand window	41 Number 3
42 SubCommand radius	42 SubCommand Zoom all	42 Number 4
43 SubCommand near	43 SubCommand Zoom extent	43 Number 5
44 SubCommand distance	44 SubCommand Zoom previous	44 Number 6
45 SubCommand set	45 SubCommand Zoom window	45 Number 7
46 Number 0	46 SubCommand radius	46 Number 8
47 Number 1	47 SubCommand Arc angle	47 Number 9
48 Number 4	48 LOW FREQUENCY WORDS	48 Number 100
49 Number 5	49 NUMBERS	49 HIGH FREQUENCY
50 OTHER	50 OTHER	50 OTHER

9.3.2 Blackboard models of system behaviour

Using the behaviour data, models of system behaviour were developed. The framework for the model and the procedures for constructing the model were described in Chapter 5. The models illustrate how and why recruitment of behavioural knowledge differed between the users of different CAD systems. A review of the data will be made in terms of the model's content and operation.

9.4 RESULTS

9.4.1 General

This section reports general results relating to the homogeneity of the sample; the effects of task order and task plan on performance.

Group equivalence

The null hypothesis states that there is no difference between the two experimental groups due to the assignment. This was confirmed by an ANOVA test ($F(1,22)=1.44$, $p>.05$) on the performance scores (a composite score of drawing errors and entities) obtained from performing the training task.

Effects of task order and task plan

A two-way ANOVA test carried out on the performance data showed no significant effect of task order on product quality ($F(1,47)=0.08$, $p>.05$); production time ($F(1,47)= 0.06$, $p>.05$); and production efficiency ($F(1,47)=0.19$, $p>.05$). In other words, the order of presentation of the task conditions was well counterbalanced. But there exists a significant effect of task plan on production costs, with $F(1,47)=10.22$, $p<.005$ for time cost and $F(1,47)=9.8$, $p<.005$ for efficiency cost. However, the effect of task plan on product quality was insignificant ($F(1,47)=1.46$, $p>.05$). This means that the design problems were not comparable in complexity but comparable in terms of the errors that might be made. The two-way interactions (Task order x Task plan), however, were not significant.

An analysis of the questionnaire (see Question 4, Appendix 25) showed that 62.5% of the subjects found both task plans to be equally difficult while 37.5% found one to be more difficult than the other. This suggests that there is a discrepancy between actual and perceived difficulty. Overall, 87.5% agreed that the tasks were interesting.

9.4.2 Effects of system on behaviour

This analysis is performed to test the predictions in Section 9.1.3. To this end, the findings from the integrated system groups (Systems D and E) will be compared with those of the unitary speech group (System C). A complete summary of the ANOVA results is given in

Appendix 26. Table 9.2 presents the group results (n=24) for some visual behaviour types.

Eye gaze to input/output devices

Within a scored duration of 480 seconds, System D subjects made, on average, 224 eye transitions (ie. one transition per 2.17 secs.), while System E subjects generated 255 eye transitions (ie. one transition per 1.97 secs.). System C subjects made 194 eye transitions (ie. one transition per 2.56 secs.). These differences in total eye gaze were highly significant, $F(2,45)=9.34$, $p<.0005$. Scheffe´ test revealed that both integrated systems differed significantly from the unitary speech system but were not different from each other. This means that, on the whole, the integrated systems incurred more eye transitions.

There were no significant differences between the systems in the duration or frequency of gazing at the graphics screen. Taking just the duration results, the time spent looking at the graphics screen are: System D = 44.6%, System E = 41.8% and System C = 40.3%. This means that all three systems are the same on this measure.

The duration of looking at the text screen did differ significantly between system subjects, $F(2,45)=9.46$, $p<.0005$. A Scheff e´ test showed System D (40.7%) to be equal to System E (32.3%) and System C (48.5%); but System E differed significantly from System C. This means that users of the integrated System E looked at the text screen for shorter periods of the time than the unitary speech users. There is a tendency, however, for System D users to gaze at the text screen for the same amount of time as System C users. It could be said that System E is a better system where text screen gazing is concerned.

In terms of the number of times subjects gazed at the text screen, there was no significant difference between systems (see Table 9.2). So, although the integrated systems incurred less time, the frequency of eyes transiting to the text screen is the same as the unitary speech input system.

These results suggest that using integrated speech-manual input systems leads to: (1) equal amount of time spent gazing at the graphics screen; and (2) shorter periods of eye gaze to the text screen. The latter behaviour is therefore more optimal and generally supports the predictions in Section 9.1.3. The results, however, did not support the predictions concerning graphics screen gazing. Of the integrated systems, System E appears to incur less behavioural costs.

The difference in gaze between the systems is also very significant in the duration and frequency of looking at the graphics tablet. The ANOVA tests revealed the differences to be

TABLE 9.2

Effects of Unitary Speech Input and Integrated Speech-Manual Input Systems on Visual Behaviour - Eye Gaze to Specific Targets (n=24)

	System C	System D	System E	ANOVA	
	mean frequency	*mean frequency*	*mean frequency*	*F(2,45)*	*p*
Eyes-gaze-All targets	2.56	2.17	1.97	9.34	.001
Eyes-gaze-Graphics screen	.16	.18	.16	1.44	.25
Eyes-gaze-Text screen	.16	.17	.15	1.16	.32
Eyes-gaze-Graphics tablet	.0	.03	.13	172.25	.000
Eyes-gaze-Keyboard	.02	.02	.01	1.57	.22
Eyes-gaze-Drawing plan	.04	.05	.05	0.62	.54
Eyes-gaze-Speech list	.01	.0	.0	0.57	.57
	% duration	*% duration*	*% duration*		
Eyes-gaze-Graphics screen	40.3	44.6	41.8	0.67	.52
Eyes-gaze-Text screen	48.5	40.7	32.3	9.46	.001
Eyes-gaze-Graphics tablet	0.2	3.9	15.4	147.43	.000
Eyes-gaze-Keyboard	2.6	1.8	1.7	0.88	.42
Eyes-gaze-Drawing plan	5.9	7.3	6.6	1.18	.32
Eyes-gaze-Speech list	0.7	0.7	0.5	.23	.80

System C=Unitary Speech Input; System D=Integrated System 1; System E=Integrated System 2

significant at .01 alpha level and greater, with F(2,45)=147.43 for duration and F(2,45) =172.25 for frequency. In both cases, Scheffe´ tests showed System E to be markedly different from Systems D and C; and System D to differ from System C as well. Taking the results of the integrated systems alone, with System E, users spent 15.4% of the time gazing at the tablet compared with 3.9% using System D. The frequency of gaze is .13 per second with System E and .03 per second with System D. This indicates that System D is the better of the two in terms of both duration and frequency of tablet gazing.

The keyboard, however, is looked at for the same amount of time by all three systems. The differences in duration and frequency of eye gaze between subjects were not significant. This indicates that the use of the integrated systems has not reduced the time spent gazing at the keyboard as well as the frequency of doing so.

The above results concerning input devices suggest that there were: (1) less frequent and shorter periods of eye gaze to the tablet by System D compared with System E; and (2) equal amount of time spent gazing at the keyboard by both systems. The findings relating to graphics tablet gazing support the predictions in 9.1.3, but the predictions concerning keyboard gazing were not supported. Despite this, of the two, it could be said that System D is a better system on these measures.

Hand manipulation of input devices

The purpose of this analysis is to test the prediction that the use of integrated systems reduces the duration and frequency of operating the manual input devices. The results will be based on the dominant hand used to manipulate the tablet. Table 9.3 summarises the group results (n=24) on specific manual behaviour types.

The duration of the hand being idle is significant between systems, F(2,45)=3.47, p<.04. But a Scheffe´ test performed on the data showed no two groups differed at p=.05, thus suggesting that all three systems kept the hand(s) equally idle. The frequency of hand being idle is highly significant, F(2,45)=100.04, p<.0001. The post-hoc test showed using System E kept the hand idle more frequently (.16) than System D (.06) or System C (.05). Systems D and C, however, did not differ. So, of the integrated systems, System E has a tendency to keep the hands less busy. This behaviour is thus more optimal.

In terms of individual hand activity, there exists no significant difference between systems in the number of times the hand spent drawing (see Table 9.3). But the differences in the duration of drawing just reached significance, F(2,45)=3.40, p<.05. This difference was not significant with Scheffe´ test, indicating that the integrated systems enabled the hand to

TABLE 9.3

Effects of Unitary Speech Input and Integrated Speech-Manual Input Systems on Manual Behaviour - Patterns of Hand Use (n=24)

	System C	System D	System E	ANOVA	
	mean *frequency*	*mean* *frequency*	*mean* *frequency*	*F(2,45)*	*p*
Hand-Idling	.05	.06	.16	100.04	.000
Hand-Drawing	.03	.03	.04	1.23	.30
Hand-Search menu item	.0	.0	.02	51.35	.000
Hand-Select command	.0	.01	.11	221.03	.000
Hand-Key-in-Keyboard	.01	.01	.0	1.51	.23
	% duration	*% duration*	*% duration*		
Hand-Idling	70.7	64.0	62.3	3.47	.04
Hand-Drawing	23.9	30.4	21.9	3.40	.04
Hand-Search menu item	.0	1.0	4.0	28.84	.000
Hand-Select command	.0	0.8	6.9	198.43	.000
Hand-Key-in-Keyboard	1.1	0.8	0.2	1.79	.18

System C=Unitary Speech Input; System D=Integrated System 1; System E=Integrated System 2

spend the same amount of time in drawing. With System D, the duration of drawing is 30.4%; with System E it is 21.9% of the time on task.

As for data entry (command and numerical data), System E subjects spent 6.9% of the time entering data via the tablet. The frequency of this occurrence is .11 per second. System D subjects, on the other hand, spent less than 1% of the time; the frequency of manipulating the tablet for this entry is also very low, .01 per second. The differences in data entry between groups were significant as shown in Table 9.3, suggesting that System E incurred more behavioural costs than Systems D or C.

The findings here suggest that the use of the integrated systems results in the hand: (1) being more frequently busy with System D than E but equally busy in terms of duration; (2) spending the same amount of time in drawing; and (3) spending more time and being more busy entering data via the tablet with System E than D. The hand idleness results did not support the predictions for System D. The findings, however, confirmed the predictions on command entry for each integrated system. Of the two, it could be said that System D is a better system, enabling the hands to draw much more than entering commands.

Verbal content of speech

This analysis is to assess the nature of speech input use, with respect to the type of utterances made and the type of speech errors incurred. Table 9.4 summarises the group results (n=24) for verbal behaviour types and Table 9.5 provides a summary of the speech content (ie. utterance type).

In terms of speech recognition, the differences between systems were highly significant, $F(2,45)=21.74$, $p<.0001$. The result of Scheffe´ test showed System E (.04) to differ significantly from System D (.09) and System C (.10). But System D is equal to System C. Therefore, it could be said that System E enabled better speech recognition than Systems D or C.

Similarly, the frequency of word repeats due to substitution, rejection and spurious errors was significantly different between systems ($F(2,45)=10.95$, $p=.0001$). A post-hoc comparison produced similar patterns as above: System E had fewer repeats (.02), while Systems D and C incurred equal number of repeats (.03 and .04, respectively). Further analysis of the speech error data showed Systems D and E producing the same number of substitution errors as System C. The frequency of verbal repeat due to this error type is .02 per second with System D and .01 per second with System E. The differences between systems were also very significant (see Table 9.4).

TABLE 9.4

Effects of Unitary Speech Input and Integrated Speech-Manual Systems on Verbal Behaviour - Content of Speech (n=24)

	System C *mean* *frequency*	**System D** *mean* *frequency*	**System E** *mean* *frequency*	**ANOVA** *F(2,45)*	*p*
Word-Recognition	.10	.09	.04	21.74	.000
Word-Repetition (total)	.04	.03	.02	10.95	.001
Word-Repeat-Substitution	.04	.02	.01	14.89	.000
Word-Repeat-Rejection	.02	.02	.01	2.37	.11
Word-Repeat-Forgetting	.0	.005	.006	6.01	.005

System C=Unitary Speech Input; System D=Integrated System 1; System E=Integrated System 2

TABLE 9.5

Group Results (n=24): Verbal Content of Speech

Utterance type	Unitary Speech System C	Integrated System D	Integrated System E
	% frequency	*% frequency*	*% frequency*
Command			
Recognised	46.96	52.07	24.56
Repeat	17.14	17.06	9.92
SubCommand			
Recognised	19.90	19.82	12.71
Repeat	7.51	6.42	1.44
Numeral			
Recognised	3.72	0.37	14.64
Repeat	0.74	-	4.65
Basic words			
Recognised	3.02	2.56	13.73
Repeat	0.74	1.43	6.20
Retrain word	0.27	0.28	12.15

From Table 9.5, it is evident that Systems D and C incurred more verbal repeats of commands plus subcommands than System E, but not of numerals and basic words. However, it is interesting to note that the latter involved more retraining of the word (12.15%) than Systems D or C (0.28% and 0.27%, respectively). Because high frequency commands were on the tablet for System E, this has led to reduced use of the speech input, which in turn led to increased retraining of words.

An important finding concerning the integrated systems is the high usage of commands that were not available within the system. System E users verbalised a number of *high frequency* commands (8.8%), while System D users verbalised a number of *low frequency* commands (2.8%). This tendency to forget that some of the commands were in different modes led to some repeat errors. An ANOVA test performed on repeat errors due to forgetting showed $F(2,45)=6.01$, $p<.005$ to be very significant. The integrated systems did not differ markedly with each other; the frequency of such errors with Systems D and E is .01 per second. This suggests that the assignment of data to device mode in these systems was inflexible.

The findings here have shown that the use of integrated systems led to: (1) better speech recognition, especially with System E; (2) fewer repeats due to substitution errors by both systems; (3) increased retraining of words due to low usage by System E; and (4) some forgetting errors in both systems. The data supported the predictions for System E, but not for System D. Overall, it could be said that System E is a better system than System D in terms of speech input.

9.4.3 Effects of system on performance

This analysis is to ascertain the extent to which the use of integrated systems affected performance. Table 9.6 presents the ANOVA tests on product quality and production costs (time and efficiency). It is evident from Table 9.6 that there were no significant differences between Systems C, D and E on all three measures of performance. This means that the use of the integrated systems to support task performance would produce the same outcome (product quality and costs) as the unitary speech system. This finding did not support the prediction in Section 9.1.3. Comparatively, there is reason to believe that the present configuration of the unitary speech system has resulted in a reduction of production costs (efficiency) than the previous system configuration of System B (Chapter 8).

9.4.4 Correlation of behaviour and performance

This analysis is to examine the relationship between behaviour and performance variables for integrated systems. Table 9.7 gives a summary of significant correlation results, for one-tailed probability tests. Only the main results will be presented. (Note: product quality refers to the

TABLE 9.6

Effects of Unitary Speech Input and Integrated Speech-Manual Input Systems on Performance - Product Quality and Production Costs

	System C	System D	System E	ANOVA	
	mean	*mean*	*mean*	*F(2,45)*	*p*
Product quality	6.88	7.00	8.67	0.62	.54
Production cost (time)	42.19	44.04	37.99	0.62	.54
Production cost (efficiency)	0.97	0.98	1.33	2.73	.08

System C=Unitary Speech Input; System D=Integrated System 1; System E=Integrated System 2

TABLE 9.7

Correlation of Behaviour and Performance Measures - Experiment 2 results

System D : Integrated Speech-Manual Input		
Behaviour with Performance variable	*r (12)*	*p*
Frequency of eye gaze to:		
Graphics tablet and product quality	-.54	.03
Graphics tablet and production cost (time)	-.55	.03
Keyboard and production cost (time)	.65	.01
Keyboard and production cost (efficiency)	.54	.04
Duration of eye gaze to:		
Graphics tablet and product quality	-.51	.05
Graphics tablet and production cost (time)	-.51	.05
Frequency of hand:		
Drawing and product quality	-.63	.01
Idling and production cost (efficiency)	.66	.01
Pressing function key and production cost (time)	.87	.00
Pressing function key and production cost (efficiency)	.74	.003
Duration of hand:		
Idling and product quality	.54	.04
Idling and production cost (time)	.52	.04
Pressing function key and production cost (time)	.82	.001
Pressing function key and production cost (efficiency)	.65	.01
Frequency of word:		
Recognition and product quality	-.64	.01
Repetition and product quality	.66	.01

System E : Integrated Speech-Manual Input		
Behaviour with Performance variable	*r (12)*	*p*
Frequency of eye gaze to:		
Text screen and production cost (efficiency)	.59	.02
Graphics tablet and production cost (efficiency)	.66	.009
Speech list and production cost (time)	.72	.004
Speech list and production cost (efficiency)	.70	.006
Duration of eye gaze to:		
Graphics screen and production cost (time)	-.53	.04
Graphics screen and production cost (efficiency)	-.75	.003
Text screen and production cost (efficiency)	.68	.007
Keyboard and product quality	.50	.05
Speech list and production cost (time)	.83	.00
Speech list and production cost (efficiency)	.88	.00
Frequency of hand:		
Idling and product quality	.73	.004
Drawing and production cost (time)	-.66	.009
Drawing and production cost (efficiency)	-.57	.03
Entering command and product quality	.60	.02
Duration of hand:		
Drawing and production cost (efficiency)	-.73	.003
Entering command and product quality	.60	.02
Frequency of word:		
Recognition and production cost (efficiency)	.70	.006
Repetition and production cost (time)	.57	.03
Repetition and production cost (efficiency)	.87	.000

number of errors in the drawing. Thus, low product quality means high errors.)

System D

With System D, frequency of eye gaze to the graphics tablet correlates negatively with product quality. This means that the more frequently the eyes gazed at the tablet, the fewer the errors in the drawing. Also, the more frequently the hand is involved in drawing, the better is the product quality ($r(12)=-.63$, $p=.01$). This implies that high levels of drawing activity may lead to less errors. This is supported by a significant, positive correlation between duration of hand idleness and product quality ($r(12)=.54$, $p<.04$). In other words, if the hand is left idle for long periods of the time on task, there is greater tendency to create more errors.

In terms of speech input, frequency of single pass recognition correlates negatively with product quality, $r(12)=-.64$, $p<.02$. That is, an increase in word recognition will lead to fewer drawing errors. Thus, improving speech input recognition should improve the quality of the task product.

An implication of the various findings above is that using System D, the user is able to spend more time in drawing, providing there is increased speech recognition. This in turn would result in a better task product. Therefore, keeping the hand(s) more idle is detrimental to performance.

Another significant finding concerns frequent eye gaze to the keyboard. This behaviour type correlates positively with both production time and efficiency costs ($r(12)=.65$, $p=.01$ and $r(12)=.54$, $p=.03$, respectively). This means increased gazing at the keyboard will increase the time it takes to generate a drawing entity, and the amount of data required to produce an entity. In short, the outcome is neither time nor data efficient. (For other significant results, see Table 9.7.)

System E

Like System D, frequency of hand idleness correlates positively with product quality, $r(12)=.73$, $p=.004$. This means the more frequently the hand is left idle, the tendency to make errors is increased. Also, errors are increased if the hand spends more time entering commands via the tablet. The correlations between duration and frequency of command entry with product quality were significant and positive (see Table 9.7). This emphasises the importance of reducing command entry via the tablet so as to increase the quality of the drawing.

With System E, the longer the time is spent gazing at the graphics screen, the less time and data are needed in entity-generation ($r(12)=-.54$, $p<.04$ for time and $r(12)=-.75$, $p=.003$ for data). The importance of increasing graphics screen gazing to enhance performance is further emphasised here. This is supported, in part, by a positive correlation between the frequency and duration of text screen gazing with production cost (efficiency) (see Table 9.7). Thus, the more frequently and the longer the periods of eye gaze at the text screen, the less data efficient is performance.

High frequency of eye gaze to the hardcopy speech list correlates significantly with high production costs ($r(12)=.72$, $p=.004$ for time and $r(12)=.70$, $p=.006$ for efficiency). In other words, the use of an offline speech list as a performance aid could slow down performance as it incurred more time and data to produce a line, etc. Gazing frequently at the tablet menu could also increase the amount of data per entity, thus causing performance to be less data efficient.

Other important correlation coefficients are the same as those found with System D, namely, the correlations between high frequency and duration of drawing with low production costs; and high verbal repeat with high production cost (time) (see Table 9.7).

In sum, the use of System D implies that the quality of the task output may be affected by reduced drawing activity and poor speech recognition. Increasing the use of the keyboard (as a backup facility to spoken commands) may incur more production costs. The use of System E, on the other hand, may affect the task output if there was increased command entry via the tablet and/or increased keyboard use. The use of a hardcopy speech list tends to increase production costs. Lastly, this particular system stressed the importance of increasing graphics screen gazing and reducing text screen gazing in order to improve performance.

9.4.5 Other questionnaire findings

This section presents some important findings relating to the problems experienced by subjects in using the systems to perform the tasks. Users' preferences for systems will also be examined.

Problems in using integrated systems

Subjects identified the following problems (listed in order of frequency of occurrence, given in brackets):

- *remembering*: difficult to remember what had to be spoken and which had to be input via the tablet (16)
- *constraint*: allocation of some words to speech and some to tablet constrained fluency in carrying out the task (13)
- *recall*: problem of recalling location of some words in tablet menu (13)
- *confusion*: problem of identifying command-to-device mapping led to confusion (7)

- *recall*: difficult to recall some spoken words (7)
- *performance*: device-combination slowed performance (5)
- *coordination*: difficult to coordinate speaking and pointing activities (4)
- *delay*: long lags between each drawing operation due to recall problem (3)
- *vocabulary size*: unmanageable for speech (2)

To overcome the problems, subjects used a number of coping strategies, such as: practise through trial and error; be patient and persistent; refer to speech list/manual; look out for cues from the system (eg. no response); use one mode (eg. tablet) continuously; minimise speech use; keep to well-learned commands; think and decide slowly; repeat drawing procedures; increase familiarity with commands and errors.

System preference

In response to question 3.3.1 (see Appendix 25), 13 of the subjects preferred the dual-mode (ie. speech-plus-tablet) while 11 preferred the single mode (unitary speech). Those who did not prefer the integrated systems gave the above problems as their reasons. Those who preferred them claimed that it was more interesting and a novelty given that the tablet was easy to use and reliable, while speech input enabled greater attention to the screen. They added that in bimodal systems, one mode could serve as a backup to the other. However, some claimed that the problems might outweigh the advantages (eg. less head movements to the tablet, thus less strain on the neck, etc.).

When queried on which unitary input mode they preferred (ie. comparing the unitary tablet mode in the training session with the unitary speech mode in the experimental session), 15 preferred the tablet mode and 7 the speech mode, while 2 did not prefer either modes. Those who preferred the unitary manual system gave problems with speech recognition as reasons for their tablet preference. However, they would prefer speech input if the system was more reliable. E2S24 explained that using unitary manual system, he has control over performance but with speech input, control lies with the device.

9.4.6 Conclusion

On the basis of the above findings, it could be concluded that both integrated systems are better in supporting CAD tasks than the unitary use of speech input, particularly in reducing the time spent in gazing at the text screen. System E especially is effective in reducing word repetition. In addition, the problems documented in Experiment 1 (relating to keyboard gazing and data inefficiency) were resolved through this hybrid design. But the inflexibility of the approach has incurred some forgetting errors on the users' part. Also, the use of performance aids such as a hardcopy speech list has negative effects, that is, it increases performance costs. Because both hybrid systems are equally effective in supporting the task and have their strengths and limitations, it could be said that both systems are suitable alternatives to unitary speech and/or manual input systems.

9.5 DISCUSSION

This section discusses the central findings in terms of the system behaviour model. The discussion will be in four parts. The first part reviews the findings in terms of the model. The second part summarises the findings by comparing the strengths and limitations of the integrated systems with the unitary speech system. The third part discusses the problems documented with the integrated systems and identifies possible solutions to the problems. The last part presents some design guidelines derived from this experiment for use in the next experiment.

9.5.1 Comparisons of system behaviour models: unitary speech input versus integrated speech-manual input systems

As indicated in Chapter 5, the number of levels in the model would be the same between systems. There is reason to believe that the models would differ in terms of: (1) the identity of KSs; (2) the levels of KS operation; and (3) the amount of KS recruitment. Although all three systems involved the use of both input modes - speech and manual - there would be differences in the types of knowledge recruited. In particular, the integrated systems used more generative Tool Management KSs, in the form of Tablet-KSs and Hand KSs, at the Action and Movement levels than the unitary speech system. The latter, on the other hand, recruited more Speech KSs rather than Hand KSs at the Action level, and more Task Specific KSs, especially commands and numerics.

Another difference would be in the amount of KS recruitment, which is controlled by the knowledge executor (scheduler). The use of unitary speech input required greater recruitment of Text screen and Speech knowledge than the integrated systems. On the bases of user skill and a history of KS use (ie. duration and frequency of use), the scheduler orders the trigger of these KSs. Because commands were mainly spoken, this simplifies the task of the scheduler: it is able to identify particular KSs that are crucial to support task performance, although the more frequent and/or longer recruitment of such KSs (eg. Text screen KS) might not result in optimal behaviour. With the integrated systems, the scheduler would tend to rely more on the users' skill in manipulating the input devices. Therefore, the allocation of resources (eg. Eyes KSs) will depend on the ability to operate the tablet without visual monitoring during command/data entry. Because there is a lack of this skill, for System E users, in particular, the scheduler allocates more time in order for users to search and select the menu items from the tablet menu. Thus, resulting in non-optimal behaviour.

The model has helped to understand the triggering of particular behaviours during CAD performance, and as in Experiment 1, differences between the systems could be related to the model's content and operation. Further differences in knowledge recruitment between systems are summarised in the next section.

9.5.2 Assessment of strengths and limitations

In light of the comparisons between integrated speech-manual systems and unitary speech system (Section 9.4), the following discussion focuses on the advantages and disadvantages of the integrated systems only.

Behaviour assessment

In terms of the model, the *advantages* of using Systems D and E (versus System C) are:

(1) with both systems, there were equal recruitment of Eyes-Graphics screen KSs and Eyes-Keyboard KSs in both duration and frequency. Given that both systems did not differ from System C on these behaviour types, there is reason to believe that the size and composition of the design vocabulary may have an effect on the behaviours. Increasing graphics screen gazing and reducing keyboard gazing would account for better performance (see Section 9.4.4).

(2) with both systems, the recruitment of Eyes KSs to the text screen is for shorter periods of the time on task and less frequent too. However, there is a tendency for System D to recruit Eyes-Text screen KSs for a longer duration than System E. This suggests that, of the two, System E is better at reducing text screen gazing. Since a reduction in this behaviour type would improve performance (Section 9.4.4), it becomes necessary to reduce this further in future hybrid designs.

(3) with both systems, there were shorter periods of Hand KSs recruitment for manipulating the input devices. This meant that the dominant hand was less busy for a significant proportion of the time during CAD performance. System D, however, tended to keep the hand more frequently busy compared with System E. Given the positive correlation between hand idleness and product quality, the need to minimise hand use, in particular command entry via the tablet, is therefore crucial in bimodal systems.

(4) with both systems, the recruitment of Hand KSs for drawing is the same in terms of duration and frequency. It should be noted, however, that increased hand use may affect the quality of the task output. This supports the point made earlier.

(5) with System D, there was reduced recruitment of Eyes-Graphics tablet KSs given that the high frequency commands were spoken. This reduction of off-screen eye transitions to the tablet is considered optimal.

(6) with System E, the use of Voice KSs for repeating commands occurs less frequently. Speech recognition is far better with this system than with Systems D or C.

The main *disadvantages* of using Systems D and E are:

(1) with System E, there was considerable recruitment of Eyes and Hand KSs to the tablet menu, in both duration and frequency. This is because the high frequency commands were on the tablet. As such, System E users spent more time in visual search as well as selecting data from the tablet menu than System D users.

(2) with both systems, there was equal recruitment of Speech KSs for reverbalising data caused by device (substitution) and user (forgetting) errors. The latter errors, due largely to the design strategy, implies that the present design was inflexible. Hence, both systems required the user to remember the input mode for different data type. This may overload user memory.

Performance assessment

Both systems are equal in producing a task output at low production costs (time and data efficiency) and with high quality. In short, the systems are time and data efficient besides being error deficient.

To summarise, the use of the integrated systems has resolved the problem of text screen gazing incurred by the unitary speech input system. Because each system is different in its configuration, System D is better at resolving the problem of off-screen transitions to the tablet incurred by unitary manual input and/or System E. The latter however is better at resolving the problem of speech confusability incurred by unitary speech input. It could be said that system behaviour is much improved by these hybrid systems. However, there is still a need to increase further on-screen gazing to the graphics screen, while simultaneously reducing off-screen gazing to the keyboard. Until these are resolved, the present design is not optimal.

Comparisons between the models of the three systems have identified minor discrepancies between prediction and observation. It is envisaged that the discrepancies may be resolved through better design of the hybrid systems. Because users recruit behavioural knowledge differently to perform the tasks, it is crucial that future systems should be more flexible and less loading on user memory. The next section looks at some of these problems.

9.5.3 Problems documented and possible solutions

This section will highlight the problems with the integrated systems and discuss possible solutions to the problems.

Inflexible hybrid design

There are two related issues here. First, the rather inflexible allocation approach, that is, a certain proportion of the data set being allocated exclusively to each input mode, introduced some forgetting errors on the users' part. Although these errors are low (3.2% with System D and 8.8% with System E), the rigidness of the approach could generate problems for casual users because of the need to remember which input mode to use for which data.

This relates to a second issue, that is, users' limited capacity. The use of CAD systems should not overtax users' memory nor incur additional workload (hand use and visual attention) to the existing load associated with the task. Although the memory problem could be alleviated through more experience with the systems, it is important to consider users' initial difficulties in learning a CAD system. Ideally, this should not pose a problem. As a result of this inflexibility, users are split in their preferences for the speech-plus-tablet systems: only 13 of the users preferred them. Those who did not prefer the integrated systems explained that the inflexible design constrained their fluency in carrying out the task and the problem of identifying command-to-device mapping led to confusion (see Section 9.4.5).

Data entry mode and behaviour/performance tradeoffs

The correlation results indicate that the use of each system incurs a tradeoff between performance costs, particularly between accuracy and speed plus efficiency. So, using System D will result in increased drawing errors due to frequent repeats, while using System E will result in increased production costs due to frequent gazing at the tablet menu and considerable data entry by the hand. Neither situation is ideal, thus a solution needs to be found which could minimise both problems - repetition and off-screen eye transitions.

Between-screen and on-screen transitions

The use of System D led to long periods of eye gaze to the text screen although the frequency of this transition is greatly reduced. As explained in Chapter 8, the need to look at the text screen is primarily for system feedback, in particular, word recognition, error messages and prompts. It was also shown here that the need to increase further graphics screen gazing is necessary to the reduction of performance costs. This would mean reducing further text screen gazing so as to increase graphics screen gazing. One way of overcoming this problem is to distribute the amount of information available on each screen.

Because of the many problems encountered, subjects suggested some ways of improving the integrated systems (see question 3.3.3, Appendix 25). It should be pointed out that some of the suggestions tend to contradict each other and may not be implementable. They are:

- numbers should be on the tablet;
- replicate commands for both input modes;
- classify commands according to some useful index (eg. importance, frequency);
- improve speech recognition;
- allow user to choose whether to speak or use the tablet;
- clear separation of commands for speech and tablet;
- allocate speech for less important tasks.

The above suggests that solutions need to be found to improve further the design of this hybrid system. Possible solutions include: (1) increasing the options for command entry mode so that spoken commands will be provided with backup entry facilities; (2) integrating system prompts within the graphics screen so that between-screen eye transitions will be further reduced; and (3) increasing the number of templates per word so that the probability of each being correctly recognised will be increased. The benefits of each of these solutions will be considered in detail in Chapter 10.

9.5.4 Design guidelines derived

As in Experiment 1, this investigation has derived the following guidelines. (The guidelines will be numbered consecutively from the previous experiment.) These guidelines are based on the empirical findings and took account of the problems encountered by the subjects.

<u>Guideline 5</u> - *performance aids should be provided to ease memory load.*

This is because speech input involves recall from memory and in the event of high speech confusability, and the inflexible allocation strategy, memory failures increase. With System E, the use of a hardcopy speech list as an aid incurs performance costs to the user (Section 9.4.4). Therefore, a performance aid in the form of screen menus might serve this purpose.

<u>Guideline 6</u> - *input devices should be allocated flexibly to the data type.*

This is because of the inflexible hybrid designs which introduced some forgetting errors on the users' part. Both Systems D and E users tended to verbalise some commands that were not available within the particular input mode. Thus, a flexible design may be achieved by replicating commands in both input modes so that users have the choice of using either mode. This means commands should not be split between input modes.

<u>Guideline 7</u> - *display of feedback information should be allocated flexibly between screens in a dual-screen configuration.*

This is due to long periods of eye gaze to the text screen by System D users. Since the subjects were inexperienced in CAD, the tendency to rely on system prompts is increased. Because the prompts and other feedback information were mainly allocated to the text screen, this led to between-screen transitions. Therefore, providing prompts on both screens will enable the user to choose which screen to access for the information (and hence is flexible), and this in itself might help to reduce between-screen transitions.

Guideline 8 - *the number of templates per design vocabulary item should be more than one.* This is in view of the high number of word repeats incurred by System D. Although System E was better in resolving this, a single pass training as recommended for this connected speech recogniser is deemed inadequate to cope with current speech problems. Therefore, increasing the number of templates per word, as often suggested in the speech literature, might help to resolve this.

9.5.5 Conclusion

This experiment has demonstrated that integrating speech and manual input within a single system has its benefits as well as problems. Both integrated systems were successful in resolving the behavioural problems associated with unitary speech system. Hence, they could be seen as a potential solution to unitary speech or manual input systems. To increase further on-screen gazing, certain modifications of the hybrid approach need to be made. The guidelines derived from this investigation will be the basis for modifying the integrated systems.

9.6 SUMMARY

This investigation has addressed the problems of unitary speech input through integrating speech and manual input. Two alternative configurations were tested and have proven to improve behaviour and performance. Because each system has more strengths than limitations, modifying the system might help to redress some of the system behavioural problems that were documented. Improving the design of the system based on guidelines derived from this investigation should produce a better and flexible system - one which offers the user alternatives for data entry, in addition to performance aids, without overloading user's I/O resources.

NEXT CHAPTER HIGHLIGHTS

In Chapter 10, a final experiment to investigate the solutions to the problems of sub-optimal behaviour, that were documented in the present hybrid systems, will be reported. The experiment will use the same behaviour and performance measures for determining the potential of modified versions of the integrated system.

CHAPTER 10

Experiment 3: Assessment of Integrated Speech-Manual Input System as a Better Hybrid Solution - Comparisons of Performance Aids and Feedback Mechanisms

Overview

10.1 Introduction

- 10.1.1 The solution and strategy - modifying the hybrid design: the use of performance aids and feedback mechanisms
- 10.1.2 Experimental aims and predictions

10.2 Method

- 10.2.1 Location and equipment
- 10.2.2 Subjects
- 10.2.3 CAD tasks
- 10.2.4 Experimental design
- 10.2.5 Procedure

10.3 Data analysis

- 10.3.1 Scoring of behaviour and performance
- 10.3.2 Blackboard models of system behaviour

10.4 Results

- 10.4.1 General
- 10.4.2 Effects of system on behaviour
- 10.4.3 Effects of system on performance
- 10.4.4 Correlation of behaviour and performance
- 10.4.5 Other questionnaire findings
- 10.4.6 Conclusion

10.5 Discussion

- 10.5.1 Comparisons of system behaviour models: old hybrid versus new hybrid speech-manual input systems
- 10.5.2 General assessment of modified hybrid systems
- 10.5.3 Design guidelines derived
- 10.5.4 Conclusion

10.6 Summary

Next chapter highlights

CHAPTER 10

Experiment 3: Assessment of Integrated Speech-Manual Input System as a Better Hybrid Solution - Comparisons of Performance Aids and Feedback Mechanisms

OVERVIEW

This chapter presents a final experiment, aimed at investigating the effectiveness of alternative configurations of integrated systems as solutions to the problems of sub-optimal behaviour observed with the integrated systems of Experiment 2 (Chapter 9). The solutions were based on guidelines derived from Experiment 2, namely, the provision of prompts within the graphics and text screens as feedback mechanisms, and the use of screen or tablet menus as performance aids. This experiment demonstrates that both modified versions of the hybrid systems are better than the old versions in optimising behaviour and supporting performance. Additionally, the novel systems helped to enhance speech recogniser performance.

10.1 INTRODUCTION

The system behavioural problems documented in Experiment 2 pertain mainly to: (1) some forgetting errors due to inflexible hybrid design; (2) some eye transitions to the text screen due to the need for command feedback; (3) eye transitions to the tablet due to visual search of tablet items for data input; and (4) verbal repeats due to speech confusability. Because these behavioural problems were shown to correlate significantly with performance measures, it is important to try to alleviate them. Therefore, potential solutions should consider two things: first, to reduce information load for both user and device (eg. vocabulary size, information display); second, to support user learning and preference (see Smith & Mosier, 1986). The next section discusses the likely solutions.

10.1.1 The solution and strategy - modifying the hybrid design: the use of performance aids and feedback mechanisms

The hybrid systems may be modified in two significant ways: (1) providing performance aids to reduce memory load and to support user preference; and (2) providing feedback mechanisms to support user learning and guidance.

Bailey (1982) defines a *performance aid* as a device or document containing information that a person uses to complete an activity. It should not be confused with training aids which are devices (eg. manual) designed to promote the *learning* of a particular skill, for future use. In contrast, performance aids are designed to assist the performance of tasks after the skills have been learned. The most significant benefits of performance aids (Bailey, 1982, p. 447) are:

(1) reduction of errors (the user relies less on long-term memory);
(2) increased speed of certain task performance (reduced uncertainty can lead to faster responses);
(3) reduced training requirements (although users must be taught to use the performance aid, they do not have to learn and remember all the information contained in the aid); and
(4) lowered minimum selection requirements (in many situations, a well-designed performance aid allows the work to be done by a person with fewer skills and less knowledge).

In terms of the thesis, (1) and (2) above relate to performance, while (3) and (4) are concerned with behaviour.

There are two types of performance aids that are used frequently. The first kind assists in the memory of specific items of information (eg. a written shopping list to use at the grocery). The second kind provides step-by-step guidance for performing an activity or executing a set of procedures (eg. step-by-step assembly instructions which accompany an unassembled computer system).

Therefore, a performance aid, as used here, is an aspect of the device containing information that will be used <u>during</u> performance. It helps by reducing the cognitive processing requirements of a CAD task, particularly by reducing the amount of information to be remembered.

In light of the problems documented in Experiment 2, there is a need for such aids in hybrid systems. This is due to: (1) the division of data between input modes which tends to impose an additional load on user's working memory; and (2) the nature of speech input which involves recall from memory, in addition to the confusability aspects of speech which tend to confuse the user. Thus, Guideline 5 states that speech input should be provided with performance aids. The purposes are to assist remembering, and most importantly, to be used as a backup facility when speech recognition drastically fails. This is in line with suggestions on alternative entries for speech input (eg. Smith and Mosier, 1986; Waterworth & Talbot, 1987).

Using a hardcopy command list as an aid was found to incur some behavioural costs for the user, which in turn was found to correlate significantly with performance (see Chapter 9). The use of screen menus as a performance aid might be more advantageous because they increase graphics screen gazing. A screen menu is a listing of textual items (in this case, commands) that are displayed online. Screen menus can be displayed permanently (ie. visible at all times) on screen, popped-up or pulled-down via pointing. An alternative aid is a tablet menu like the one used in System A (Experiment 1) which groups the commands together in a tablet overlay.

Guideline 6 states that the input devices should be allocated flexibly to the data type. As explained in Chapter 9, a flexible system is one that provides the user with options in making responses. The rigidity of the old hybrid approach resulted in forgetting errors on the users' part. (The term *old* is used here to refer to the previous hybrid design of Experiment 2.) This is because the whole command set was divided between input modes into two distinct and non-overlapping sets. Keeping the commands together, but separating them from the numerical data, should help to resolve the forgetting problem. Also, users might benefit from flexibility in inputting commands via two modes - spoken and tablet menu or spoken and screen menu. In this experiment, command entry via speech will be a default data entry mode, while those via tablet or screen menus will be backup modes.

Guideline 7 states that feedback information should be allocated flexibly between screens in a dual-screen configuration. The design of information display in dual-screen systems is crucial to promote effective interaction between user and computer. The problem of scarce screen space is often exacerbated by the need to accommodate menus, prompts and other control objects on screen (Newman & Sproull, 1979). Therefore, decisions need to be made about what information to display and where, bearing in mind user requirements. In Chapter 3, the importance of feedback in task performance was discussed. From Experiments 1 and 2, it was learnt that users required feedback from the command interpreting process, in particular from prompts and system messages located on the text screen.

A *prompt* is a message output by the system to indicate that an input is required from the user (Coats & Vlaeminke, 1987). The user's reply to the prompt may invoke a particular task process (eg. looking at the drawing) or activity by the dialogue process (such as the provision of assistance in responding to the prompt) or it may supply data values to a task process. There are a number of possible formats for displaying prompts in human-computer dialogues (see Newman & Sproull, 1979). In CAD, a prompt is an output message termed a *command* prompt, displayed at the command line (Raker & Rice, 1985).

Because of users' limited knowledge and experience in CAD, there is a tendency to rely heavily on prompts which were displayed primarily on the text screen. Thus, allocating prompts to both screens should help to reduce frequent shifts in eye gaze. Also, users might benefit from this flexibility in feedback display. This strategy is in line with standard guidelines on information display (eg. Cole et al., 1987; Davis & Swezey, 1983).

Guideline 8 suggests that the number of templates per word/phrase should be more than one. Most connected speech recognisers suggest a single pass training (see Chapter 7). This has been found to be inadequate in the context of this research. Given the high number of repeats, coupled with the occasional need to retrain the recogniser, it is essential to improve device training by providing two templates per vocabulary item. The retraining facility in this recognition system is only available on the text screen; its frequent use would increase between-screen transitions. Since subjects will be provided with an alternative mode for command reentry (ie. the backup facility), minimising the use of the retraining facility should help to reduce the time spent gazing at the text screen. Also, it was learnt from both experiments that subjects found retraining to be time-consuming (ie. there was a high behavioural cost attached to retraining).

In sum, the modifications to the integrated systems are: (1) increasing the options for command entry mode so that spoken commands will be provided with backup facilities; (2) incorporating prompts within the graphics screen so that information will be distributed between screens; and (3) increasing the number of templates per word so that the probability of each being recognised will be increased. In line with these modifications, two alternative configurations of speech-plus-manual input system were derived. To provide a baseline condition against which to compare the performance of the new hybrid systems, the old hybrid system (similar to System D) will be employed. (Note that the old system was inflexible, and hence this new version differs slightly.)

The following variables are kept constant in all three systems:

- default command entry via speech recogniser
- graphical and numerical input via graphics tablet
- text entry via keyboard
- graphical and system status information on graphics screen
- textual information and command line on text screen.

Two variables will be manipulated in the experimental design, namely, input modes for backup commands and prompts for feedback mechanisms. The three configurations were:

- System F (old hybrid): tablet menu (backup commands) and prompts on text screen.
- System G (new hybrid 1): screen menu (backup commands) and prompts on text screen.
- System H (new hybrid 2): tablet menu (backup commands) and prompts on graphics plus text screens.

10.1.2 Experimental aims and predictions

The objectives of this experiment are two-fold:

(1) to investigate the potential of the modified hybrid systems as solutions to the problems

of sub-optimal behaviour observed in Experiment 2; and

(2) to document the nature of the problems in using the modified hybrid versions.

As in previous experiments (ie. Experiments 1 and 2), assumptions must be made about the relationships between system components and behaviour/performance. The predictions are expressed in IF-THEN statements. Those variables that are kept constant throughout the conditions (see above) will be excluded, as well as the following predictions concerning behaviour that are assumed to hold across systems. These take into account findings from both experiments, 1 and 2. There will be no significant differences between systems in:

(1) the frequency and duration of eye gaze to the keyboard;

(2) the frequency and duration of dominant hand manipulating the input devices for drawing and data entry; and

(3) the frequency of verbal repeat due to device and user errors.

However, behavioural differences will be expected of the following.

Taking **System F** (old hybrid) as a control condition, the disadvantages of using it over Systems G and H (new hybrids) are:

IF

TASK is to input information for conducting design activity
COMPUTER INPUT DEVICE is graphics tablet for entering tablet backup commands, and OUTPUT DEVICE is text screen for displaying system prompts
USER is able to look at the screens while speaking <u>but</u> is not able to input tablet commands without directly looking

THEN (behaviour)

- Moderate frequency and duration of eye gaze to the graphics screen.
- Moderate frequency and duration of eye gaze to the text screen.
- High frequency and duration of eye gaze to the graphics tablet.
- Moderate frequency and duration of hand manipulating the tablet for menu selection.

This would result in performance being:

- Moderate product quality.
- Moderate production costs (time and efficiency).
- Moderate user acceptability.

The outcome is that behaviour and performance will be sub-optimal.

The advantages of using **System G** (new hybrid 1) relative to System F are:

IF

TASK is to input information for conducting design activity
COMPUTER INPUT DEVICE is graphics tablet for entering screen backup commands, and OUTPUT DEVICE is text screen for displaying system prompts
USER is able to look at the screens while speaking <u>but</u> is not able to enter screen menu commands without directly looking

THEN (behaviour)

- Increase in frequency and duration of eye gaze to the graphics screen.
- Decrease in frequency and duration of eye gaze to the text screen.
- Decrease in frequency and duration of eye gaze to the graphics tablet.
- Decrease in frequency and duration of hand manipulating the tablet for menu selection.

This would result in performance being:

- High product quality.
- Low production costs (time and efficiency).
- High user acceptability.

The outcome is that behaviour and performance will be significantly improved over System F.

The advantages of using **System H** (new hybrid 2) over System F are:

IF
TASK is to input information for conducting design activity
COMPUTER INPUT DEVICE is graphics tablet for entering tablet backup commands, and OUTPUT DEVICES are graphics and text screens for displaying system prompts
USER is able to look at the screens while speaking but is not able to enter tablet menu commands without directly looking

THEN (behaviour)

- Increase in frequency and duration of eye gaze to the graphics screen.
- Decrease in frequency and duration of eye gaze to the text screen.
- Decrease in frequency and duration of eye gaze to the graphics tablet.
- Decrease in frequency and duration of hand manipulating the tablet for menu selection.

This would result in performance being:

- High product quality.
- Low production costs (time and effciency).
- High user acceptability.

The outcome is that behaviour and performance will be significantly improved over System F.

For the new hybrids, there will be a tradeoff in performance costs. However, both are expected to improve behaviour, leading to enhanced performance.

10.2 METHOD

10.2.1 Location and equipment

The experimental setting, demonstrator system and recording equipment for this experiment are those used in Experiments 1 and 2. The essential changes are in the design of the screen and tablet menu overlays to suit the requirements of this experiment. The screen menu was displayed on the right-hand side of the graphics screen (default location). It was designed to be two-level: the first level contained major commands, the second level contained subcommands, arranged alphabetically. To select the second level of commands, subjects were required to point to *next menu*, and to return to the first level, they had to select *last menu* on the second level. The tablet menu was designed to be similar to the screen menu (see Appendix 27).

10.2.2 Subjects

Unlike the previous two experiments, subjects were paid £4.00 for their participation. Sixteen subjects aged between 18 and 46 years (mean= 28 years) took part in this 4.5 hour experiment. Nine of the subjects were male and 7 were female; 11 were British, 5 were non-British. Half of the sample had normal vision, the other half wore glasses or contact lenses. One subject was left-handed, another was ambidextrous, the rest were right-handed. The sample was equally divided into students and non-students. The minimum academic qualification of the subjects was A-level.

Three of the subjects had no computer experience, the rest had some (<6 months) or a lot (>6 months) of experience. Six subjects had some CAD experience (the novices), the remaining ten had none at all (naive users). Four of the novices had participated in Experiment 2 (for the first time), one in the optimisation study, and one had not been in any of the experiments. This means 5 of the novices have had prior experience with the test system, but each only once. Each experimental condition had the same number of novices and naive subjects. The subjects will be referred to by their identity number (eg. E3S6).

10.2.3 CAD tasks

The tasks were based on representative CAD tasks. They involved draughting bathroom layouts based on drawing plans provided. The plans showed detailed arrangement of bathroom ware, based on plans from a design book by Niesewand (1986). Three plans were derived; the first plan (Appendix 28a) was used in the training session, the second and third plans (see Appendix 28b) were employed in the experimental sessions. The same precautions taken in Experiment 2 were applied here to ensure that both experimental plans contained the same number of drawing entities and objects.

The speech vocabulary list (Appendix 29) was the same for all three systems, comprising 48 commands (major, subcommands and basic words). The backup commands, however, had only 40 commands (8 basic words were excluded).

10.2.4 Experimental design

Like Experiment 2, the design was a oneway mixed ANOVA design with all subjects tested in the control condition; one half in the first experimental condition and the other half in the second experimental condition. The independent variable was type of system; the dependent variables were behaviour and performance measures. Subjects were assigned to the experimental conditions using their performance scores obtained in the training session and their CAD experience. The order of task presentation was also counterbalanced across conditions and task plans.

Behaviour and performance measures

This experiment used the following metrics:

(1) behaviour measures: frequency and duration of eye gaze to I/O devices, plan and speech list; frequency and duration of hand manipulation of graphics tablet and keyboard; and frequency of single pass recognitions and verbal repeats due to device and user errors as verbal content of speech; and

(2) performance measures: product quality, production costs and user acceptability.

10.2.5 Procedure

The experiment was conducted in two sessions: a training session (2.5 hours) and an experimental session (2 hours), each separated by between 1-2 days.

Training session

Each subject completed four phases of the training:

Phase 1. *Introduction to the experiment and CAD system* (15 minutes). The procedure for conducting this session was the same as in previous experiments. The training materials (introduction and general instructions) are given in Appendix 30a.

Phase 2. *Train on speech recogniser* (30 minutes). Subjects were first trained on how to use the recogniser (user and device training). This was followed by template training of the speech commands. Words with a score greater than 8 were retrained and the check procedure was repeated for these words.

Phase 3. *Learn AutoCAD using the recogniser* (60 minutes). Subjects were trained in AutoCAD using the recogniser for command entry, the tablet for numeric entry and the keyboard for text. This is to enable positive transfer of learning from the training to the experimental sessions. No backup facilities were provided and the prompts were on the text screen.

Phase 4. *Practise doing a CAD task* (45 minutes). This practice task was performed using the same input devices as in the demonstration session. Subjects were allowed to ask whenever in doubt or to refer to the manual whenever necessary. The time to complete the task was set at 45 minutes. Each drawing was assessed to obtain a performance score based on the number of errors and drawing entities.

Experimental session

This session was conducted in two phases.

Phase 1. *Practise* (20 minutes). Subjects were first trained on how to use the backup commands in the tablet or screen menu. They then practised using speech commands. Any problems in speech recognition were resolved during this trial.

Phase 2. *Perform two draughting tasks and complete two questionnaires* (95 minutes). Each subject performed two tasks. The time allowed for each task was set at 35 minutes. The subjects' behaviour were recorded on video for a duration of 15 minutes per task. The same recording procedure as in Experiments 1 and 2 was repeated here. The recording form used in both sessions is given in Appendix 30b.

The task instructions were displayed on cards. Each condition had separate instructions (see Appendix 31). Subjects read the instructions prior to the commencement of each task. The instructions reminded them: (1) of the backup and feedback facilities available for the task; (2) that they should use the backup facility whenever their spoken commands (default) were not recognised; (3) that they should check for prompts from the relevant screen; and (4) that they should work accurately and as efficiently as possible. In addition, subjects were also instructed to remain as consistent as possible in their speech production.

At the end of each task, subjects were required to complete a short questionnaire (see Appendix 32). This questionnaire is the source for information relating to subjective ratings of the system, user satisfaction and performance, problems experienced with each system, strategy for overcoming the problems, system preference and recogniser rating.

10.3 DATA ANALYSIS

10.3.1 Scoring of behaviour and performance

The videotapes provided the behaviour data; the drawings and questionnaire provided the performance data (objective and subjective measures, respectively). The manual behaviours scored were the same as in Experiment 2 (see Table 9.1, Chapter 9). Table 10.1 summarises the menus for scoring visual behaviour and verbal content of speech. Using VITAS (see Chapter 4), a 12-minute segment was selected after 2 minutes of the tape had elapsed and this was scored continuously. Steps were taken to identify the behaviour types prior to scoring. Each drawing was scored following the same scoring procedure of previous experiments. The questionnaire was coded and processed using SPSS/PC+.

Using parametric tests, a oneway ANOVA was performed to test the effect of systems on the various measures. To test if any two groups differed significantly on the measures, the means were tested with a Scheffe′ test. As explained in Experiment 2, this procedure is conservative; any differences between the main effects are therefore real differences. To determine the relationship between behaviour and performance variables, Pearson correlation

TABLE 10.1

Categories of Visual and Verbal Behaviours

Visual Behaviour

Direction of Looking
1 Text screen
2 Graphics screen
3 Tablet menu
4 Screen menu
5 Graphics prompts
6 Tablet/Stylus
7 Keyboard
8 Drawing Plan
9 Command list
10 Experimenter
11 Elsewhere

Verbal Behaviour

Utterance type
1 SR command
2 CAD command
3 DOS command
4 CAD SubCommand

Recognition Status
1 Single pass recognition
2 Recogniser error (Substitution)
3 Recogniser error (Rejection)
4 User error (Repetition) recognised
5 User error (Repetition) not recognised

Command/Data list
1 SR Command Listen
2 SR Command Goodbye
3 SR Command Hello
4 SR Command Switch off
5 DOS Command Cancel
6 DOS Command Enter
7 DOS Command Return
8 CAD Command Arc
9 CAD Command Break
10 CAD Command Circle
11 CAD Command Copy
12 CAD Command Erase
13 CAD Command Fillet
14 CAD Command Grid
15 CAD Command Insert
16 CAD Command Line
17 CAD Command Move
18 CAD Command Oops
19 CAD Command Ortho
20 CAD Command Pan
21 CAD Command Redraw
22 CAD Command Snap
23 CAD Command Text
24 CAD Command Trace
25 CAD Command Zoom
26 CAD Subcommand ?/Help
27 CAD SubCommand 2-point
28 CAD SubCommand 3-point
29 CAD SubCommand aligned
30 CAD SubCommand all
31 CAD SubCommand angle
32 CAD SubCommand centre
33 CAD SubCommand close
34 CAD SubCommand drag
35 CAD SubCommand ellipse
36 CAD SubCommand endpoint
37 CAD SubCommand last
38 CAD SubCommand near
39 CAD SubCommand off
40 CAD SubCommand on
41 CAD SubCommand previous
42 CAD SubCommand radius
43 CAD SubCommand rectang
44 CAD SubCommand undo
45 CAD SubCommand window
46 OTHER

tests were performed on the data. All predictions were tested at the 5% alpha level.

10.3.2 Blackboard models of system behaviour

As before, Blackboard models of system behaviour were developed from the behaviour data (see Chapter 5 for the framework). Only the Action KSs will be reported here.

10.4 RESULTS

10.4.1 General

This section presents the results on the homogeneity of the sample, the effects of task order and task plan on performance.

Group equivalence

An ANOVA test performed on the performance scores obtained in the training session revealed $F(1,14)=0.35$, $p>.05$ to be not significant. This means the sample was homogeneous prior to being assigned to the experimental groups.

Effects of task order and task plan

A two-way ANOVA test carried out on the performance data showed no significant effects of task order and of task plan on performance. The results for task order are $F(1,20)=3.80$, $p>.05$ (product quality) and $F(1,20)=1.79$, $p>.05$ (production time), suggesting that the tasks were well counterbalanced across conditions. The effect of task plan on product quality is $F(1,20)=1.63$, $p>.05$ and on production time is $F(1,20)=0.12$, $p>.05$. Unlike Experiment 2, the plans here were found to be comparable in complexity (ie. the number and type of objects and errors each contained).

The two-way interactions of Task order x Task plan for the error data were significant ($F(1,20)=4.73$, $p<.05$), suggesting that the errors differed at different levels of the plan and task. The Task order x Task plan interactions for the production time data were not significant ($F(1,20)=2.01$, $p>.05$). An analysis of the questionnaire (see Appendix 32) showed that 9 of the subjects found one task to be more difficult than the other, while 7 found both tasks to be equally difficult. The differences in user skill probably account for the significant interaction between task order and plan.

10.4.2 Effects of system on behaviour

This analysis is to test the predictions in Section 10.1.2, specifically the effects of system on behaviour. A complete summary of the ANOVA results appears in Appendix 33. Table 10.2 presents the group results (n=16) of a few, select visual behaviour types.

TABLE 10.2

Effects of Old Hybrid and New Hybrid Systems on Visual Behaviour - Eye Gaze to Specific Targets (n=16)

	System F	System G	System H	ANOVA	
	mean frequency	*mean frequency*	*mean frequency*	*F(2,29)*	*p*
Eyes-gaze-All targets	2.45	2.57	2.30	0.74	.48
Eyes-gaze-Graphics screen	.18	.18	.23	4.64	.02
Eyes-gaze-Text screen	.15	.15	.12	2.29	.12
Eyes-gaze-Graphics tablet	.04	.01	.03	10.29	.001
Eyes-gaze-Keyboard	.0	.0	.001	0.77	.47
Eyes-gaze-Drawing plan	.05	.04	.06	1.10	.35
Eyes-gaze-Command list	.003	.004	.003	.02	.98
	% duration	*% duration*	*% duration*		
Eyes-gaze-Graphics screen	54.6	62.0	65.2	6.16	.006
Eyes-gaze-Text screen	28.9	27.0	19.7	4.13	.03
Eyes-gaze-Graphics tablet	6.4	1.4	5.0	9.37	.001
Eyes-gaze-Keyboard	0.07	0.04	0.09	0.17	.84
Eyes-gaze-Drawing plan	8.2	7.3	8.1	0.08	.93
Eyes-gaze-Command list	1.1	0.9	0.7	0.08	.92

System F=Old Hybrid System; System G=New Hybrid System 1; System H=New Hybrid System 2

Eye gaze to input/output devices

Within a scored duration of 720 seconds, System G subjects made, on average, 285 eye transitions (ie. one transition per 2.57 secs.); System H subjects made 321 eye transitions (ie. one transition per 2.30 secs.); and System F subjects produced 307 eye transitions (ie. one transition per 2.45 secs.). The differences in total eye transitions were not significant (see Appendix 33). Generally, all three systems incurred the same number of eye transitions.

From Table 10.2, it is evident that there were highly significant differences between the systems in the duration and frequency of eye gaze to the graphics screen. The duration results showed: System G users spent 62% of the time, System H users spent 65.2% and System F users, 54.6%. An ANOVA test found $F(2,29)=6.16$, $p<.006$ to be highly significant: System H differed significantly from System F in the Scheffe´ test; System H, however, is equal to System G, and the latter is equal to System F.

The frequency results produced a similar pattern as the duration results: System H is the same as System G but different from System F, while System G is the same as System F. An ANOVA test of the results showed it to be significant, $F(2,29)= 4.64$, $p<.02$. Both duration and frequency results indicate that System H enabled longer periods and more frequent eye gaze to the graphics screen than System F. On this basis, it could be said that System H is better than System G since the latter tended to incur similar eye gaze as System F.

The duration of looking at the text screen did differ significantly between system subjects, $F(2,29)=4.13$, $p<.03$. A Scheffe´ test showed System H (19.7%) to differ significantly from System F (28.9%) but equal to System G (27%). Again, System G incurred the same amount of time as System F in text screen gazing. This indicates that users of the new hybrids spent less time looking at the text screen than the old system, but there is a tendency for System G users to spend the same duration as System F users. As expected, there exists no significant difference between different system users in the frequency of gazing at the text screen (.15, .12 and .16 per second for Systems G, H and F, respectively). Although the new hybrids incurred less time, the frequency of eye transitions to the text screen is the same as the old hybrid.

These results suggest that using the modified versions of the integrated speech-manual input system leads to: (1) frequent and longer periods of eye gaze to the graphics screen; (2) shorter periods of eye gaze to the text screen. This behaviour is therefore more optimal and supports the predictions in Section 10.1.2. Of the new hybrids, it could be said that System H is a better system since there is a tendency for System G to behave the same as the old hybrid.

The differences in the frequency and duration of gaze to the graphics tablet were highly significant between systems (F(2,29)=9.37, p<.0001 for duration and F(2,29)=10.29, p<.0001 for frequency). In both cases, Scheffe´ tests showed System G to be markedly different from Systems H and F; but System H to equal System F. In other words, the duration of gazing at the tablet is shorter with System G (1.4%) compared with System H (5%) and System F (6.4%). The frequency of tablet gaze is .01 per second with System G, .03 per second with System H, and .04 per second with System F.

The keyboard, as predicted, was gazed at equally between system users. The differences in duration and frequency of eye gaze between subjects were not significant (see Appendix 33).

The conclusions about eye gaze to input devices are: of the new hybrids, System G reduced the duration and the frequency of gazing at the tablet. But both systems incurred the same duration and frequency of gazing at the keyboard. The results support the predictions in Section 10.1.2.

Hand manipulation of input devices

This analysis is to test the prediction that the use of new hybrid systems reduces the duration and the frequency of manipulating the tablet, particularly in menu selection. The results will be based on the dominant hand used to manipulate the tablet. Table 10.3 summarises the group results (n=16) on a few, select manual behaviour types.

In terms of hand idleness, the duration and frequency of the hand being idle did not differ significantly between systems. There were also no significant differences between the systems in the number of times the hand spent drawing as well as the total time spent drawing. This means that the resources recruited for drawing are the same between systems, thus accepting the null hypothesis.

The frequency and duration of the hand inputting commands and numerical data via the tablet were not significantly different between users. Given that the input modes for data entry were kept constant across systems, this finding is as expected. But the time spent selecting the menu item to reinput the command did differ significantly between systems, F(2,29)=5.25, p=.01. Scheffe´ test performed on the duration results showed System G (5.6%) to differ greatly from Systems H (2.4%) and F (2.8%); System H is the same as System F. In other words, with System G, the duration of selecting the backup command is longer than System H. The frequency of menu selection did not differ between systems (see Table 10.3).

The findings here suggest that the use of the new hybrids results in the hand: (1) being equally busy in carrying out the task; and (2) spending the same amount of time in drawing and

TABLE 10.3

Effects of Old Hybrid and New Hybrid Systems on Manual Behaviour - Patterns of Hand Use (n=16)

	System F	System G	System H	ANOVA	
	mean frequency	*mean frequency*	*mean frequency*	*F(2,29)*	*p*
Hand-Idling	.07	.07	.07	0.05	.95
Hand-Drawing	.04	.04	.04	0.28	.76
Hand-Locate menu item	.02	.03	.01	1.86	.17
Hand-Enter command	.03	.02	.02	0.26	.77
Hand-Enter numeric	.005	.004	.003	0.56	.58
	% duration	*% duration*	*% duration*		
Hand-Idling	58.6	58.9	58.0	0.06	.94
Hand-Drawing	35.3	32.5	37.4	1.45	.25
Hand-Locate menu item	2.8	5.6	2.4	5.25	.01
Hand-Enter command	1.6	1.6	1.1	0.65	.53
Hand-Enter numeric	0.7	0.8	0.5	0.70	.50

System F=Old Hybrid System; System G=New Hybrid System 1; System H=New Hybrid System 2

entering data. But of the two, it could be said that System H is a better system than System G, enabling the hands to spend less time in selecting the backup commands via the tablet menu.

Verbal content of speech

This analysis is to assess speech recogniser performance as a function of systems. It is predicted that all three systems should be the same in recognition performance and in the type of device and user errors incurred. Table 10.4 summarises the group results (n=16) for verbal behaviour types.

Mean speech performance obtained during use, based on a single pass recognition, is as follows (Note: the results are based on relative %): System G (66.1%), System H (68.9%) and System F (64.2%). The differences between systems (based on absolute values) were not significant, $F(2,29)=0.19$, $p>.05$, hence confirming the prediction. The findings are however encouraging: System H, in particular, improved recognition from the usual low of 62-64% to a moderate level of 69%. Unlike previous experiments, individual results showed much better recognition, in the moderate threshold region of 70% to 75%.

The frequency of recogniser errors (substitution, rejection, spurious) was not significantly different between systems ($F(2,29)=0.24$, $p>.05$). All three systems incurred the same number of device errors as predicted. Likewise, the frequency of user errors (forgetting, commission errors, wrong input, etc.) was also insignificant ($F(2,29)=0.09$, $p>.05$). This finding indicates that the new hybrid design is flexible.

10.4.3 Effects of system on performance

To determine the extent to which the use of the new hybrid systems affects performance, this section reports the results of ANOVA tests on product quality, production costs and user acceptability (see Table 10.5). There were no significant differences between Systems F, G and H on all measures of performance, thus the predictions were not supported. This could be due to the size of the manipulation of the independent variables. Compared with Experiment 2, this manipulation is relatively minimal, involving only the performance aids and feedback facilities to support task. Furthermore, speech input was held constant across systems, thus its effect on performance was similar between systems. The results, however, confirm further that the hybrid system design is better in supporting task performance than the unitary input systems.

Given that subjects used the speech recogniser as a default entry mode, their ratings of the device did not differ significantly. Subjects rated the recogniser similarly ($F(2,29)=0.01$, $p>0.05$). The group mean ratings ranged between 59.6% to 60.6%. This finding supported the

TABLE 10.4

Effects of Old Hybrid and New Hybrid Systems on Verbal Behaviour - Content of Speech (n=16)

	System F	System G	System H	ANOVA	
	mean frequency	*mean frequency*	*mean frequency*	*F(2,29)*	*p*
Word-Recognition	.09	.09	.10	0.19	.83
Word-Repeat-Device error	.03	.03	.03	0.24	.79
Word-Repeat-User error	.01	.01	.01	0.09	.92

System F=Old Hybrid System; System G=New Hybrid System 1; System H=New Hybrid System 2

TABLE 10.5

Effects of Old Hybrid and New Hybrid Systems on Performance - Product Quality, Production Costs and User Acceptability

	System F	System G	System H	ANOVA	
	mean	*mean*	*mean*	*F(2,29)*	*p*
Product quality	8.81	9.13	9.75	0.06	.94
Production cost (time)	45.84	44.54	46.84	0.04	.96
Production cost (efficiency)	1.76	1.70	1.79	0.08	.93
User acceptance (performance)	62.64	59.18	58.97	0.17	.85
User acceptance (satisfaction)	58.40	50.22	65.60	1.18	.32
User acceptance (recogniser rating)	60.57	59.62	60.27	0.01	.99

System F=Old Hybrid System; System G=New Hybrid System 1; System H=New Hybrid System 2

verbal behaviour results on recognition performance.

10.4.4 Correlation of behaviour and performance

The purpose of this analysis was to map the relationship between behaviour and performance variables for the modified integrated systems. Table 10.6 gives a summary of significant correlation results for one-tailed probability tests. Only a few, select results will be presented.

System G

With System G, high duration of eye gaze to graphics screen correlates significantly with high production efficiency, $r(8)=.66$, $p<.04$. This means the more time is spent gazing at the graphics screen, the more data are required per entity drawn. The use of a screen menu for backup commands helps to increase graphics screen gazing (and hence is optimal), but the time involved in menu selection is increased as well. This is probably due to the sensitivity of the stylus which did not ease the selection process.

The correlations between the duration of hand drawing and production costs were very significant ($r(8)=-.68$, $p<.04$ for time, and $r(8)=-.69$, $p<.03$ for data efficiency). Both results suggest that production costs are much reduced with increased time spent on drawing. Also, subjects tended to rate the recogniser high with increased drawing time. The correlation between the variables was significant, $r(8)=.73$, $p=.02$. (Other significant results are shown in Table 10.6.)

System H

With System H, high frequency and high duration of eye gaze to the graphics screen correlate significantly with low production time (see Table 10.6). This means the more time is spent in graphics screen gazing, the less time is needed to produce an entity. The correlations between duration of eye gaze to graphics screen and subjects' ratings on performance and satisfaction were also positive. This indicates that subjects tended to rate their performance and satisfaction high with increased time spent looking at the graphics screen. On the contrary, production time increases with increased time spent gazing at the text screen. This tends to result in low ratings on performance and satisfaction (Table 10.6). The results here confirmed the importance of increasing graphics screen gazing and reducing text screen gazing which this system has managed to accomplish.

The frequency and duration of eye gaze to the tablet correlated significantly with users' ratings of the recogniser, their performance and satisfaction. The more time is spent gazing at the tablet, the lower is the rating by subjects on all three performance measures (Table 10.6). This confirmed that frequent and long periods of tablet gazing are not approved by subjects,

TABLE 10.6

Correlation of Behaviour and Performance Measures - Experiment 3 results

System G : New Hybrid Speech-Manual Input		
Behaviour with Performance variable	*r (8)*	*p*
Frequency of eye gaze to:		
Text screen and user acceptability (performance)	-.65	.04
Drawing plan and production cost (efficiency)	-.77	.01
Duration of eye gaze to:		
Graphics screen and production cost (efficiency)	.66	.04
Graphics screen and user acceptability (satisfaction)	-.69	.03
Drawing plan and production cost (efficiency)	-.73	.02
Drawing plan user acceptability (performance)	.91	.001
Drawing plan and user acceptability (satisfaction)	.98	.000
Duration of hand:		
Drawing and production cost (time)	-.68	.03
Drawing and production cost (efficiency)	-.69	.03
Drawing and user acceptability (recogniser rating)	.73	.02
System H : New Hybrid Speech-Manual Input		
Behaviour with Performance variable	*r (8)*	*p*
Frequency of eye gaze to:		
Graphics screen and production cost (time)	-.82	.006
Text screen and user acceptability (performance)	-.88	.002
Text screen and user acceptability (satisfaction)	-.78	.01
Graphics tablet and user acceptability (performance)	-.70	.03
Graphics tablet and user acceptability (satisfaction)	-.74	.02
Graphics tablet and user acceptability (recogniser rating)	-.71	.03
Command list and production cost (time)	.75	.02
Duration of eye gaze to:		
Graphics screen and production cost (time)	-.70	.03
Graphics screen and user acceptability (performance)	.68	.03
Graphics screen and user acceptability (satisfaction)	.70	.03
Text screen and production cost (time)	.80	.008
Text screen and user acceptability (performance)	-.76	.01
Text screen and user acceptability (satisfaction)	-.70	.03
Graphics tablet and user acceptability (performance)	-.76	.01
Graphics tablet and user acceptability (satisfaction)	-.79	.009
Graphics tablet and user acceptability (recogniser rating)	-.73	.02
Command list and production cost (time)	.75	.02
Frequency of hand:		
Idling and product quality	.74	.02
Entering command and product quality	.65	.04
Selecting menu and user acceptability (performance)	-.75	.02
Selecting menu and user acceptability (satisfaction)	-.74	.02
Duration of hand:		
Selecting menu and production cost (time)	.88	.002
Selecting menu and user acceptability (performance)	-.75	.02
Selecting menu and user acceptability (satisfaction)	-.69	.03
Entering command and product quality	.64	.04
Entering command and user acceptability (performance)	-.70	.03
Entering command and user acceptability (satisfaction)	-.70	.03
Frequency of word:		
Recognition and production cost (time)	-.65	.04
Repeat (speech errors) and user acceptability (performance)	-.65	.04
Repeat (speech errors) and user acceptability (satisfaction)	-.72	.03

thus resulting in poor ratings of the system on these measures.

However, errors are much reduced if the hand spends longer duration and more frequently selecting commands via the tablet menu. The positive correlations between frequency and duration of command entry with product quality were significant ($r(8)=.65$, $p<.04$ and $r(8)=.64$, $p<.05$, respectively). But users' ratings on performance, satisfaction and recogniser deteriorated with high frequency and longer durations of command reentry. Similarly, selecting the menu items more frequently and the longer periods of time spent on the selection process correlated significantly with all three user acceptability measures (see Table 10.6). These findings imply the dissatisfaction associated with increased menu entries. (For other significant findings, see Table 10.6.)

In sum, it could be said that using System G, users are able to spend more time on drawing, leading to a reduction in production costs. But there is a potential of spending longer durations on screen menu selection. System H, on the other hand, enables the user to spend a significant proportion of the time gazing at the graphics screen. This led to a reduction in production costs, and an increase in user acceptability. However, there is the potential of incurring performance costs if the tablet menu is used more frequently and for longer durations in selecting menu items.

10.4.5 Other questionnaire findings

This analysis is to complement the findings on behaviour and performance. An analysis of the problems experienced in using the systems, system preference and speech input assessment will be made.

Task performance

Subjects identified the following problems in using the systems to carry out the tasks: speech recognition; knowing the command syntax; limited experience in CAD; remembering how to do; eye-hand coordination; discomfort with headset microphone; and background noise. To overcome some of these problems, subjects named the following strategies: increase practice through redrawing; request help (verbally) from the experimenter; be patient; and focus on one screen (ie. graphics screen).

To carry out the tasks, 6 subjects traded off speed for accuracy. Nine tried to be efficient (ie. accurate and quick), but one subject (E3S9) claimed that it was better to be quick and not necessarily accurate. This subject did make the most drawing errors in both tasks. Their emphasis on accuracy probably explains the low errors in the drawing.

Hybrid system preference

Subjects were evenly split in their preference for the system. Within the System G group (n=8), 4 ranked System G as first preference and System F as second, and vice versa. Similarly, within the System H users (n=8), 4 preferred using System H followed by System F, and vice versa. This means that users preferred both the new and the old systems. Reasons for their preferences include: rely less on memory; less commands to remember; increased attention on screen while the hand draws; more variety in input operations, thus reducing monotony; and less confusion as commands are available in both modes.

Speech input assessment

Since speech input was the default command entry mode, separate ANOVA tests performed on users' ratings of the recogniser showed no significant differences between the new hybrid systems. Mean ratings on variables - tiring, enjoyable, confusing, and easy to use - were the same between system users (see Appendix 32). However, there was a significant difference in the mean ratings on speed ($F(1,14)=5.46$, $p<.04$). System G users tended to rate the recogniser as quicker than System H users.

Despite the clear instructions not to repeat a command, instead to use the backup facility, subjects still repeated because:

- quicker than looking for the command in the menu
- testing the recogniser's performance
- habit
- effects of training
- forgot the availability of the backup facility
- overcoming frustration
- like/enjoy speaking.

The above findings suggest that despite the constraints of speech input, users still preferred it as a primary input mode.

10.4.6 Conclusion

The various behaviour findings can be summed as follows: both modified versions of the integrated systems are better in supporting CAD tasks than the old hybrid system, especially in increasing the duration and frequency of eye gaze to the graphics screen, and reducing the time spent gazing at the text screen. System G reduced the frequency and duration of tablet gazing while System H reduced the duration spent selecting menu items. In terms of performance, both hybrid systems produced similar effects on product quality, production costs and user acceptability. The correlation results confirmed previous findings on potential performance problems of increased eye gaze to the text screen and graphics tablet.

In conclusion, the new hybrid design is flexible and has resolved the system behavioural problems documented in Experiment 2. Both systems are equal in performance and were equally preferred by the novices and naive users.

10.5 DISCUSSION

The principal findings of this experiment will be discussed in three parts. The first part reviews the findings in terms of the model; the second part summarises the significant findings concerning the new hybrids. Problems documented with each system will be highlighted. The third part summarises the guidelines derived from this experiment.

10.5.1 Comparisons of system behaviour models: old hybrid versus new hybrid speech-manual input systems

Given that all three systems used speech input as the default mode for command input, and the tablet as a device for selecting backup commands (tablet menu or screen menu), the recruitment of Task Specific KSs at the SubTask level, and Tool Management KSs at the Action and Movement levels is similar between systems. The main differences are in terms of the identity of the individual KSs and the length of each KS use. For example, the old hybrid system recruited more Text screen KSs than the new hybrids. The latter used more Graphics screen KSs than the former. The new hybrids, on the other hand, differed from each other in terms of the knowledge available on the graphics screen. With System G, the knowledge relates to the availability of screen commands as a backup facility; with System H, the knowledge is concerned with the availability of prompts as a feedback mechanism.

The operations of the model are determined, in part, by the knowledge executor. Because the hybrid systems are more flexible, the functions of the knowledge executor in ordering the KSs are simplified further. There is a tendency for it to base its judgement on the importance of the KSs to optimal behaviour, as well as on the history of KS use. By applying different control algorithms for the triggering of particular KSs, it is able to protect the user against errors, in addition to reducing production costs, thus resulting in better performance.

In sum, it could be said the models have contributed toward understanding better the recruitment of knowledge using the modified versions of the hybrid systems. In particular, they have helped to identify critical user behaviours during CAD performance which must be considered in configuring different components of the system.

10.5.2 General assessment of modified hybrid systems

This experiment has produced two modified versions of the integrated systems, using design guidelines derived from Experiments 1 and 2. The novel systems have proven to enhance behaviour and performance. System G which uses a screen menu as performance aid helped to increase graphics screen gazing while reducing text screen gazing. System H which incorporates system prompts on both graphics and text screens as feedback mechanism helped to minimise further text screen gazing while increasing graphics screen gazing.

Both hybrid systems are also equal in performance in that they both: produced the same number of drawing errors; incurred the same amount of time in entity generation; utilised the same amount of data per entity drawn; were accepted by users as being efficient; as such were equally preferred.

In terms of the model, System G users recruited more manual resources to select the menu, while System H users recruited more visual resources to locate and select the menu. This means that the use of different systems involves a tradeoff between different types of behavioural knowledge. With System G, the recruitment of Hand KSs for selecting commands from the screen menu is for longer periods of the time. This is partly due to: (1) the sensitivity of the stylus which was previously claimed as being difficult to manipulate; and (2) the proximity of the commands in a screen menu. As a result of this, subjects had to spend more time manipulating the stylus in menu selection.

With System H, the recruitment of Eyes KSs occurs more frequently and for longer periods of time in order to search and guide the selection of data from the tablet menu. This deployment of resources is a function of device design. The use of an off-screen input device, such as the tablet, which splits the information display into two parts, necessitates eye transitions to operate the device visually.

On the bases of both behaviour and performance indices, two conclusions could be made: first, both systems are flexible, and second, both systems are effective in supporting CAD tasks. Clearly, the new hybrids are better solutions to the problems of sub-optimal behaviour observed with the old hybrid design. Discrepancies between normative and performative models of system behaviour were resolved through this solution. From the performative models, it was learnt that users required different behavioural knowledge to perform the task, as discussed above.

The remainder of this discussion will examine the role of the backup and feedback facilities within each system.

Role of performance aids in hybrid systems

Independent of the systems in use, all subjects, with the exception of E3S14, agreed that having backup command facility to aid speech input was useful. Some of the reasons given were it: saves tedious repetition; provides immediate and correct feedback on input; enables one to get on with the task; facilitates interaction with the system; is quicker to use than to repeat. E3S14 found the backup command facility (ie. tablet menu) not useful because it was "confusing - hard to find commands in menu". This is probably due to the order in which the commands were arranged.

Within the System G users (n=8), subjects were evenly split in their preference for the screen or tablet mode as performance aids. Only 2 subjects preferred alternative backup modes than the ones they experienced, 5 were contented with both modes and 1 was unsure. Within the System H users (n=8), 5 preferred a different mode to the tablet which they used throughout the experiments: 3 suggested the keyboard as an alternative, 2 proposed the puck. Three of the subjects who preferred the tablet mode for backup commands explained that it was more efficient as one does not have to shift the hand between different types of input devices.

The role of feedback mechanisms in hybrid systems

Irrespective of the systems, all subjects read the prompts: 2 read them all the time, 14 only sometimes. In addition to the prompts, 11 of the subjects sometimes checked the text screen for command recognition, while 5 checked this all the time. Within the System H users (who were provided with system prompts on both graphics and text screens), all claimed that they read the prompts on the text screen more than the graphics screen. But the behaviour findings found the contrary, indicating that users' perceptions differ from reality.

Amongst the System H users, the subjects were evenly split in their preference for the feedback facility. Those who found it not helpful explained that it was confusing, forgot that it existed, and were apparently conditioned to text screen gazing by virtue of the training. The same subjects also tended to not prefer having other systems information (eg. error messages) on both screens. Those who preferred more information identified speech recognition status and error messages as important feedback information.

Within the System G users, 6 subjects found the availability of system feedback only on the text screen as satisfactory. E3S8 said "you could see your past mistakes". Others explained that it was quicker to reference the text screen or that they did not use the information much. Five of the subjects did not prefer to have other systems information on both screens, while 3 preferred otherwise. The reasons for not preferring it were mainly due to potential clutter of the graphics display, and increased division of attention between displays.

10.5.3 Design guidelines derived

The guidelines derived from this experiment, numbered consecutively from Experiment 2, are:

Guideline 9 - *when speech input is used as a default entry mode, provide backup facilities, one of which should be an online facility.*

The need for a backup facility for speech input is emphasised in the literature (eg. Smith & Mosier, 1986; Waterworth & Talbot, 1987; Hapeshi & Jones, 1988). The facility should be simple to use so that the costs of use would be much less than the costs of re-verbalising the entry. This experiment suggests the use of an online facility (eg. pull-down screen menu) as a backup, in addition to an off-screen performance aid (eg. tablet menu, keyboard). The importance of increasing graphics screen gazing implies that the online facility should be allocated to this screen.

Guideline 10 - *speech input should be provided with more than one form of feedback, one of which should be prompts.*

The need for prompts to aid speech performance is discussed in the literature (eg. Williges et al., 1986). Prompts are required especially by naive users and novices in view of their limited experience with the system. This experiment suggests the use of visual prompts as a feedback facility, in addition to auditory prompts (as proposed by subjects). Other feedback mechanisms identified include error messages and status messages concerning speech recognition and data entry. Given the importance of increasing graphics screen gazing, the visual prompts should be displayed on the graphics screen.

Guideline 11 - *when speech input is the primary mode for data entry, provide alternatives for critical entries so that if the system cannot recognise an entry then another entry can be substituted.*

Because speech recognition is affected by normal variations in a user's 'task' voice, by changes in the acoustic environment and by displacement of the microphone, etc., a spoken entry that was accepted during training might not be accepted during actual use. Thus, for important entries a user should be able to use an alternative word that is acoustically dissimilar.

Guideline 12 - *the design of the tablet should be improved so that it remained unresponsive to slight pressure from the transducer.*

Sensitivity of the tablet to slight pressure from the stylus has been a problem to some users. Coupled with stylus sensitivity, this has led to numerous data reentries. Because of varying skill level and work style, users are not able to use the transducer in the optimal position suggested. This problem was also raised by Davis and Swezey (1983).

Guideline 13 - *the tablet and screen menus should be designed such that the items are not*

cluttered nor overload user memory.
Organisation of menu items is crucial to menu selection process. Items should be arranged following some logical grouping principles which do not conflict with each other nor with task. Examples of such principles are given in the literature (eg. Cole et al., 1987; McKenzie, 1988). Arrangement of tablet items in alphabetical order was detrimental to visual search but this did not affect screen menu items.

In addition to the above, there are other guidelines relating to training, error correction, etc. which are developed post hoc to the research. These will be described in Chapter 11.

10.5.4 Conclusion

The new hybrid systems have been shown to be effective in optimising behaviour and performance. The use of performance aids and feedback mechanisms that are task-relevant has helped to resolve the problem of sub-optimal behaviour observed with the old hybrid system, thus resulting in a better fit between the normative and performative models of system behaviour. The problems associated with each system may be resolved through better design of the user interface, taking into account the various user requirements. The guidelines derived from this experiment will form part of the human factors guidelines to be developed in Chapter 11, for configuring and/or designing flexible speech-plus-manual CAD systems.

10.6 SUMMARY

This experiment clearly demonstrates the significance of integrating speech and manual input within a single system. Using design guidelines derived from the previous experiments, two modified versions of the hybrid system (which provided performance aids and feedback facilities) were configured. Both systems have resolved the behavioural problems documented in Experiment 2 by increasing the duration of eye gaze to the graphics screen, reducing the duration of eye transitions to the text screen, and improving speech recogniser performance. Since both hybrids are similar in their effects on behaviour and performance, they could serve as potentially useful demonstrator CAD systems. Alternative configurations to these may benefit from the human factors guidelines derived.

NEXT CHAPTER HIGHLIGHTS

Chapter 11 describes the development of human factors guidelines for integrating speech and manual input in CAD systems, using material derived from the three experiments. To determine their usability, the proposed guidelines will be validated by system designers and their suggestions will be the basis for revising the guidelines.

CHAPTER 11

Developing and Validating Human Factors Guidelines for Integrating Speech and Manual Input in CAD Systems

Overview

11.1 Introduction

11.2 Derivation and development of guidelines

11.2.1 Guidelines application

11.2.2 Guidelines organisation

11.3 Validation of guidelines

11.3.1 Identifying the system designers

11.4 Analysis and discussion

11.4.1 Design practice and the use of guidelines

11.4.2 Content-validation of proposed guidelines

11.4.3 Usability of guidelines as design tool

11.4.4 Suggestions on guidelines refinement

11.4.5 Conclusion

11.5 Summary

Next chapter highlights

CHAPTER 11

Developing and Validating Human Factors Guidelines for Integrating Speech and Manual Input in CAD Systems

OVERVIEW

This chapter describes the development and validation of guidelines for the combined use of speech and manual input in CAD systems. The guidelines are based on the findings from the three experimental investigations of human factors solutions to the problems of non-optimal behaviour and performance. The proposed guidelines are validated by system designers to determine their clarity and usability in configuring integrated speech-manual CAD systems. Given that the guidelines needed refinement, their effectiveness as a design tool could not be determined. However, it was shown that the guidelines are potentially useful for use in configuring or designing speech-plus-manual input systems.

11.1 INTRODUCTION

There are three important aspects in the generation of human factors guidelines. The first aspect concerns their origin and development. Ideally, guidelines should have some empirical support. The second aspect relates to the refinement of the guidelines through expert validation. Ideally, guidelines should be validated by different classes of guideline users in order to ensure that they are clear and usable. The third aspect concerns the scope of their application. Ideally, guidelines should be generic in order to be applied to designing general-purpose systems.

The approach taken here is to develop the guidelines using the experimental data, to validate them using system designers with expertise in speech and/or CAD systems, and to revise the proposed guidelines following suggestions by the designers. This revision, however, will be undertaken as further work, beyond the scope of the thesis. In terms of application, the guidelines are confined to the design of multimodal CAD systems, particularly in suggesting what components of the system need to be considered in system design. Because the guidelines are derived from the experiments, there is no clear guarantee as to their effectiveness.

This chapter is in three parts. The first part describes the guidelines and their development, the second part presents the validation method, and the third part discusses the major findings from the validation process.

11.2 DERIVATION AND DEVELOPMENT OF GUIDELINES

The successful application of speech and manual input in human-computer interfaces depends not only upon technological advances in speech and/or manual input hardware, but also on the development of empirically based human factors guidelines. The term *guidelines*, as used here, refers to a set of recommendations or suggestions. Hence, they are intended to be prescriptive but without being able to offer guarantees that optimal system behaviour will result.

While great advances in speech recognition technology may be envisaged, there still remains the need to understand and maximise the effectiveness of speech interfaces using current technology. Integrating speech and manual input within a single system has proven to result in more optimal design behaviour and task performance (see Chapters 9 and 10). The guidelines proposed here are intended to improve the user interface design of CAD systems that integrate speech and manual input.

Users of CAD systems interact with a computer in order to accomplish design tasks (see Chapters 2 and 3). The users will differ in ability, training, experience and attitudes. Design of speech-manual interfaces must take into account these human factors. The investigations reported in Chapters 7 to 10 have identified problems experienced by users in using the demonstrator CAD system. These problems emerged as user complaints, indicating considerable effort on the users' part to adapt to the configured system and to develop short-term coping strategies so as to overcome the problems. The guidelines proposed here take account of user requirements expressed by users of the demonstrator system.

The need for design guidelines in the development and refinement of input devices was raised in Chapter 1 (Section 1.4.3). There is general agreement that most of the design guidelines that have been suggested lack empirical support (Meister, 1988; Nickerson, 1986). The guidelines proposed here are not based solely on opinion, rather they are derived from the investigation of solutions to the problems documented concerning the use of unitary speech or unitary manual input, and subjective assessment of the use by naive and novice users (see Chapters 8-10).

In short, the guidelines proposed here are based on empirical evidence from the three experimental investigations on the problem-solution analysis and are intended to support the optimisation of system behaviour and task performance. However, they are not complete and potential users would still need to refer to other handbooks on guidelines for advice on aspects not covered here.

11.2.1 Guidelines application

The proposed guidelines are intended to be used in two ways: (1) to configure a speech-plus-manual interface for CAD; and (2) to design or improve a speech and/or manual interface for CAD systems. Here, the term *configuration* refers to putting together various system components to make an efficient whole, whilst the term *design* denotes creating a new system or *improving* an existing system. (The term *user interface* relates to all aspects of system design that affect a user's participation in information handling transactions.)

The guidelines are intended for use by three classes of users: (1) end users; (2) implementors of CAD systems; and (3) system designers. The guidelines, however, are not developed with the third group in mind. Therefore, it is not clear whether they could apply them in designing novel integrated systems. With regard to the first and second groups of users, the guidelines are tailored to meet their requirements in configuring speech-manual input CAD systems, particularly using existing systems. The aim is to optimise the use of available systems, adding speech to the user interface, in order to derive an integrated speech-manual input system. This might provide a solution to CAD users and implementors in making good use of their available systems rather than to invest in new CAD and/or speech systems.

It should also be realised that the proposed guidelines are intended for CAD applications. In other words, it is not clear whether they could be applied to other applications such as word-processing, data base management, etc. However, with appropriate modifications, and providing there is sufficient documented evidence that the rules have been consistently applied within an application, the design rules derived from these guidelines might later be used for other applications.

Therefore, the guidelines proposed here are intended to support the configuration and design of speech plus manual input devices in CAD applications.

11.2.2 Guidelines organisation

The guidelines are formal to the extent that they are expressed as explicitly as possible. To use the guidelines effectively, they must be translated into specific design rules. "It is only specifically worded design rules that can be enforced, not guidelines." (Smith & Mosier, 1986, p. 9). This is because a guideline can be translated into different design rules by users. For example, a guideline which states that system information should be distributed in dual-screen displays might be translated by one user into a design rule that specifies where various information type should appear, such as prompts on the graphics screen and command entries on the text screen. Another user might translate the same guideline into a design rule that requires the prompts to be displayed on both graphics and text screens but command entries on the text screen only.

It is also possible that a particular guideline (eg. concerning design of speech vocabulary) could conflict with another guideline (eg. concerning allocation of device functions to data type). Failure to translate the guidelines into highly prescriptive design rules can result in inconsistent design. In other words, these guidelines are offered to users as a potential resource, rather than as a contractual design standard.

There are 24 guidelines in the initial proposed set (see Appendix 34); thirteen were derived from the experiments (see Chapters 8-10), while eleven were developed post hoc to the research, using evidence from the investigations and related HCI literature. The format for expressing the guidelines follows the standard format by Smith and Mosier (1986). The guidelines are organised in a single section; within the section, the guidelines are grouped by specific functions. Under any function, there will be guidelines pertaining to related subordinate topics. The section begins with an introductory discussion of design issues relating to data entry and information display. The discussion provides some perspective for the guidelines that follow.

The guidelines themselves are numbered sequentially in order to permit convenient referencing. Each guideline has been given a short title to indicate its particular subject matter. Following its number and title, each guideline is stated as a single sentence. Guidelines are worded as simply as possible.

A stated guideline will be illustrated by one or more examples. The examples are based on the thesis findings and/or suggestions from the relevant literature. Examples that are derived from the thesis has empirical support, hence may be more effective than those that are derived from the literature which may require further research. It is important, however, to emphasise that examples are presented here only to illustrate and are not intended to limit the interpretation of guidelines. Examples are followed by comments. The comments are clarifications of a guideline to provide the reasoning behind a guideline using material from the thesis investigations. Where a comment is related in some way to other published reports, references will be made citing author(s) and date. The references are listed in the bibliography section of the thesis.

For the purposes of illustration, four examples of the proposed guidelines are given below, each taken to represent its functional grouping.

Example 1: from 1.0 Device functions and Data allocation

1 Clearly defined functions for input devices

The functions of each input device should be well rationalised and clearly distinct from each other.

Example: Assign speech input to command entry while keyboard to text entry.

Comment: Clearly defining the functions that each device supports helps to optimise the utility of each input device and to simplify its use. This view is supported by Monk (1986) and Whitefield (1986a). Thus, a device should be assigned function(s) to which it is best suited. There is evidence to suggest that the keyboard is suited to text entry, the tablet for graphical entry, the recogniser for command entry.

Example 2: from 1.1 Speech input

13 Easy error correction for speech input

Provide simple error correction procedures for speech input, so that when a spoken entry has not been correctly recognised, the user can cancel that entry and speak again.

Example: The use of an explicit CANCEL action that is not tied to other task functions might be one way of interrupting the execution of the verbalised input.

Comment: The need for some form of error correction procedures is widely discussed in the speech literature (eg. Williges et al., 1986; McCauley, 1984; Hapeshi & Jones, 1988). Error correction procedures independent of CAD functions are needed to support speech input. It has been demonstrated that users employed the CANCEL command much more than any other command in order to terminate the execution of a confused command. Because CANCEL sometimes is confused with another word, the tendency to revert to keyboard use becomes inevitable. This in turn resulted in increased eye and hand transitions, thereby incurring costs.

Example 3: from 1.2 Manual input

16 Minimal use of manual data entry

Data entry via manual input mode should be kept to a minimum so that a user can stay with one manual method of entry, and not have to shift to another.

Example: Minimise the use of two input devices that require the same output resources so as to reduce hand transitions from tablet to keyboard and vice versa.

Comment: Shifting of one hand between two input devices sharing the same modality (such as the keyboard and tablet) has been shown to incur behavioural costs. For users who are not able to touch type or key-in tablet items without visual aid, the need to gaze away from the primary display to manipulate the input devices occurs frequently, besides incurring time. This problem was also documented by Van der Heiden and Grandjean (1984), as well as Monk (1986).

Example 4: from 1.3 Information Display

21 Flexible allocation of system information to displays

Allocate system information flexibly between displays in a dual-screen configuration so that users have the choice of which display to view for the information.

Example: Assign prompts and command feedback on the graphics and text screens, while other forms of feedback are assigned to the text screen to enable users to process information selectively from the displays.

Comment: System information (such as prompts) that is needed by one group of users (eg. novices) should be displayed on both screens. But other users (eg. experienced) should be provided with alternatives to by-pass the standard user guidance facilities. Having prompts on both screens has helped to increase graphics screen gazing in CAD, besides reducing between-screen eye transitions.

For a complete set of the proposed guidelines, see Appendix 34. The next section describes the approach to validation. This activity is part of the development process illustrated in Appendix 35a.

11.3 VALIDATION OF GUIDELINES

The purpose of this validation is to ascertain whether the proposed guidelines are: (1) clear and explicit with regard to the terminology used and the statement of the guidelines; and (2) useful for the purposes for which they were intended. The first aim will provide an indication of *face validity* (ie. the extent to which the guideline states what it purports to state), and the second aim an indication of *usability* (ie. the extent to which the guidelines can be used in system configuration and/or development). Both validity and usability will be determined subjectively on the basis of expert verbal and written protocols.

The validation process involves the use of a structured interview method. The interview schedule was divided into four sections (see Appendix 35b). First, the designers described their job experiences and educational background. Second, their current design practice, including problems arising from the practice, and the use of guidelines were discussed. Third, they assessed each proposed guideline in turn, on the basis of its clarity and suitability. Lastly, the designers judged whether the guidelines were usable for the intended users.

11.3.1 Identifying the system designers

System designers are professionals with expertise in an area of system development. This term is used here in a broad sense to include system developers, human factors practitioners and design consultants who provide professional advice. For the purposes of this validation, their areas of specialisation should include the design of user-computer interfaces for CAD and/or speech systems. Given this criterion, seven system designers (D1 to D7) with the relevant expertise were interviewed. All were male, aged between 25 to 53 years. A face-to-face interview, lasting on average two hours, was conducted with D1 to D6 and their verbal protocols recorded on audiotape. D7 was 'interviewed' by mail and his responses were recorded in an interview schedule. A short introduction to the research was made so as to provide context to the guidelines.

The first designer (D1) was a professor of information engineering who also heads a design company, which is involved in the design of interactive systems, including CAD. His involvement in CAD development began in 1966. Additionally, he has been involved in development projects relating to the design of menu-based information systems and the modelling of interactive systems. His expertise in circuit design enabled him to consider the application of the guidelines to the design of user-computer interfaces for CAD systems.

The second designer (D2) was an education training manager with a well-eastablished software company. He served the company for 3.5 years; prior to this he was a lecturer in new technology. In addition to managerial tasks, his job involved designing training material for dealers training, and the syllabus content for City and Guilds courses for CAD users. Because he trained dealers on the testbed software, he contributed to the design of user-computer interface by providing feedback information directly to the software developers of his company. This enabled changes to be made to the application before its development was completed.

Designers 3 to 6 all worked as executive engineers with the Human Factors Division of a telecommunications company. They have worked there between 3-5 years (mean=3.5 years); they were all involved in the design and development of the user interfaces for the company's products; and they were all interested in speech I/O development. D3's background was in linguistics; his area of specialisation was phonetics. He was particularly involved in the design of dialogues for developing the company's banking systems. D4's job was to evaluate systems and developing user interface standards for the design of their products. D5 was a user interface designer and a trained ergonomist. He had worked on a range of products from CAD to telephone and computerised systems. D6's educational background was in psychology and computer science; his current involvement was in user interface architecture, particularly on design issues relating to device independence and task structure.

The seventh designer previously worked in the same company as D3 to D6 before assuming his present job as a research psychologist. He has been involved in the past on multimodal input systems for applications other than CAD. He now works in developing communication aids for non-speaking people.

The above description of designer experience and skills provides a source of information for appreciating the viewpoints expressed in the next section. In particular, their comments might reflect their expertise.

11.4 ANALYSIS AND DISCUSSION

To understand how designers set about designing systems, the first part of this section discusses their design practices. This might help to identify problems experienced in the design and development process, and the extent to which they use guidelines. The second part discusses their comments on the proposed guidelines as subjective assessments of the guidelines' validity and usability. The third part discusses suggestions for further refinement of the guidelines as usable design tool. Where relevant, excerpts from designers will be cited and the quotes will appear in italics.

11.4.1 Design practice and the use of guidelines

In reality, design is likely to involve a degree of iteration. The design of a system evolves throughout the design and development process, thus the system is not simply specified and built. The use of prototypes helps to identify user difficulties, thus allowing changes to be made to the system. These views of design, as discussed in Chapter 3, are supported by the following quotes:

[I] *used iterative design procedure, that is, design and test.* [D3]

[Design] *occurs incrementally (...) We must always have a system that works so that a user can try it out. I took the view that I had to have the vision what things would be like in 10 years. It took about 20 years to get (...)* [D1]

Once we developed the course we will put together some prototype training material and run the course. And then we can look if our objectives were achieved in time. [D2]

Problems encountered in the design process include:

There are questions which there is no right or wrong answer. There is no literature to guide you. [D1]

The time factor is the biggest problem to overcome. [D2]

Difficulty in trying to demonstrate concepts rather than to redesign a new system. Not able to do task analysis adds to the problem. [D6]

In attempting to use human factors guidelines in the design process, some designers are faced with constraints ranging from time, effort to availability.

I do not know if there were any guidelines at that time. I don't know what guidelines there are now. I should do. Time is always the problem. There is a definite constraint that is money. If the guidelines are available on computers, then I'll probably use them. At the moment I *probably will have to go to the book* [Smith & Mosier]. [D1]

Some designers view guidelines in terms of objectives.

The guidelines I'm aware of for designing syllabus are in terms of objectives. We measure what we hope the course would achieve in terms of stated objectives. I suppose that is a guideline in a way, to break down the course content into easily measured objectives and work towards enabling people to achieve these objectives rather than the traditional way of listing down topics. We evaluate, part of the design process, after they have done the course and discuss whether it was achievable (...) [D2]

As far as possible, designers in the telecommunication industry used human factors guidelines, particularly their own internal guidelines. There is minimal use of other sources of guidelines.

Yes, certainly we use our own human factors guidelines. Other guidelines, not really. Hopefully at the end of the day, I shall come up with my own set of guidelines on spoken dialogue design. [D3]

If the guidelines do not cover an area of an interface that we're evaluating we use other commercial available ones to give a better guess of what the interface should be. [D4]

In sum, it could be concluded that the present design and development practice of these designers is towards user-centred design, that is the intention is to make the user the central focus of the design process rather than the system (D2). The use of prototypes (D1, D2), iterative design (D3) and human factors guidelines (D3 to D6) are ways of achieving a more usable system.

11.4.2 Content-validation of proposed guidelines

This section presents designers' general comments on the guidelines. These are mostly critical comments, as required in the validation, which will contribute to redesign of the guidelines. There were also positive comments but space precludes reporting them here. (Note: There were no general comments from D1, D2 and D7.) The quotes are used to indicate the degree of acceptability of the guidelines on important design issues. A summary of matters arising from the validation process will be made at the end of this section.

1. <u>Mixed guidelines</u>

A lot of important issues have been incorporated into the guidelines when they could have been guidelines themselves. [D3]

2. <u>Suggestions for further work</u>

Many of the guidelines seem more like recommendations for future research rather than *specifically usable guidelines, that is, something concrete that one can use right now.* [D3]

3. Definitive statements

You would have to come up with definitive statements. If you are going to suggest guidelines for integrated CAD systems this is how you're going to do it but this did not come across. [D4]

4. Generality of guidelines

The guidelines must be generic enough. If you only provide specific quantas that do not define the type that they are to square it is such a waste. [D4]

If the guidelines are only for CAD systems, it is very difficult to generalise to other systems. It's also the case if the guidelines are very general, they're going to be very useless. [D5]

5. Organisation of guidelines

The general organisation of the guidelines need to be looked at again. The order in which they occur, I personally don't think this is a good way of doing it. You need a framework for your guidelines. That's what is missing. [D4]

There might be some classification of the kinds of guidelines, some only to speech, some to the text mode. Others might deal with terminology and how you decide what words to use in speech interface. [D6]

6. User model

I didn't get any feeling of something that is binding all these together. You mentioned a model of system behaviour [in the thesis abstract]. *But I didn't see there was any model binding the guidelines in the introduction.* [D6]

7. Coverage of issues

In the individual guidelines where errors occur, there didn't seem to have any notion of confirmation, recovery and repair, collapse or degradation (...) I felt that would have given me this as a guideline, a better conceptual model about how to make decisions where the guidelines didn't cover a particular case or where they're conflicting. [D6]

The whole aspect of dialogue design seems to be pushed aside in this document. [D3]

8. Standalone guidelines

These guidelines, however good they are, must be able to stand on their own. If I have to start cross-referencing, checking other references, I just won't read it. Also, you made references to the thesis and the demonstrator CAD system. The references are not relevant to the guidelines. [D4]

9. <u>Conformance</u>

For system designers to use guidelines, however bad or high level and not really helpful, you need to show a degree of conformance. In user interface, standardisation is becoming more and more and it will increase in the future. There will be commercial pressures put on companies: that you will conform and you must use these guidelines. [D4]

10. <u>Using guidelines</u>

I would rather have introduction to the guidelines than the area to which the guidelines applied. All I got was the literature review, interesting but wasn't useful. Guidelines application is a good place to start. These are the guidelines and this is how you're going to use them. [D4]

11. <u>Target users</u>

I would have thought the system designers would be the most important group to aim a set of guidelines such as these. I don't see how the guidelines can affect the end users. It is the system designers who have the final say what is going in, his decisions will be based on the guidelines that are available. [D3]

Guidelines ought to be aimed at designers to provide a scope for individual customisation. The scope for individual selection should be within constraints set out by designers such that it comes up with a usable interface. [D4]

I would think the end users might look at the guidelines whilst configuring but certainly not a prime target. As it is, you're constrained by decisions made by designers. The designers need to be the prime target for these guidelines. [D6]

I would also argue the fact that as an individual if I were to configure a system, I won't need any guidelines. I would do what I feel comfortable. [D4]

12. <u>Clarity</u>

There seems to be a considerable overlap between what you mean by data entry and information display. There is not a great distinction between the two. [D3]

You should explain more clearly what optimisation of behaviour and performance means. You talk about costs and you don't describe what the costs are. There are tradeoffs in cost. [D6]

You're using words and people interpret words. You got to make it easy to use. The guidelines have to be fairly specific and explicit in order for people to pay attention to. You got to encourage people to use them. [D4]

You should provide a glossary of terms used in the guidelines. [D4]

You 've got to try to be more positive than just provide a policing institution. You 've got to sell it as something that takes away the need to make difficult design decisions. [D6]

13. Emphasis

That's what you're offering guidelines on - How you're going to structure the user interface. And that needs to be emphasised throughout this, the use of guidelines for user interface because at the end of the day we want to design something that is usable for the people. And not for the computer. [D4]

14. Conflicting principles

There are general principles of design as opposed to individual guidelines that can be tied to any system. You want things like consistency, learnability. Often those principles conflict and you'll have to use that general knowledge to make decisions. [D6]

15. References

Do away with references. Smith and Mosier is not so much guidelines. It's a reference manual for human factors people. It doesn't necessarily offer guidelines. [D4]

To summarise, the above excerpts suggest the following:

(1) some guidelines are clear while others are less definitive and need to be refined further;

(2) the format for expressing the guidelines and the general organisation needs re-examination;

(3) the context of the guidelines needs to be made explicit in order to understand the relationship between the guidelines and the empirical data plus the model;

(4) the guidelines need to be standalone and self-explanatory; and

(5) the procedure for using the guidelines needs to be outlined clearly in introducing the guidelines.

As D4 explained, "what you're trying to do is valid but at the same time if what you've got is not quite right, it is worth knowing...If what you're doing is for a specific system, I'd rather see that you be that specific". This means revising particular guidelines, to be discussed in Section 11.4.4. The next section discusses the status of the guidelines in terms of usability.

11.4.3 Usability of guidelines as a design tool

It is generally agreed by the designers that the guidelines are usable by others, providing suggested refinements are made to particular guidelines. However, there is a tendency to avoid the issue of whether they themselves would use the guidelines.

D7 liked the approach and thought the objectives were sound. D2 claimed that "they were very useful and sensible. I can see the logic behind (...)". He believed that the proposed guidelines would be usable to software developers in its parent company. Also, he suggested "you'll benefit by speaking to the end users. The way we used it in a training situation is different from the way it is used in a productive situation [ie. real task]. This might help to illuminate specific problems that may not arise during training." D1, however, was particularly concerned with the deliverable aspect of the guidelines. That is, for the guidelines to be easily accessible by system designers, they should be available in computerised form (ie. online), in addition to a hardcopy form.

However, there is a lack of agreement on certain usability issues, particularly by designers 3 to 6. The factors contributing to the disagreement may be summarised as follows:

(1) A problem of definition

From the comments, it is not clear whether the term 'guidelines' has a single meaning. There is reason to believe that D3-D6 are arguing for design standards (see Chapter 1) rather than guidelines *per se*, as intended. The emphasis on conformance and standardisation (see point 9, Section 11.4.2) suggests that this was the case. This caused D4 and D5 to regard guidelines by Smith and Mosier (1986) as reference material which "may not necessarily be guidelines". However, D4 qualified it, "when I'm writing guidelines like what I'm doing now, I go to Smith & Mosier. But I don't say to people go and have a look at Smith & Mosier."

(2) A problem of target usership

The guidelines were intended for three classes of users (see Section 11.2.1) with greater emphasis on the end users and CAD implementors than system designers. This was rejected by D3-D6 as it is not clear if the end users would use them. Again, there is reason to believe that the term 'configuration' is interpreted differently by the designers to mean design *per se*. Thus, it was emphasised that system designers should be prime target for these guidelines since they "make decisions for the end users".

(3) A problem of generality and coverage

The guidelines were intended for CAD applications and their derivations were based on the experimental data. The former meant that the guidelines may not be generalisable to other application domains. The latter meant that several design issues which were not investigated in the thesis could not be covered in the guidelines. These constraints, which tended to delimit the usability of the guidelines, were not fully appreciated by the designers.

The above differences in intent and meaning may account for the lack of agreement on the usability aspects of the guidelines. It should be pointed out, however, that these differences are purely on technicalities. Also, D3 to D6 have a tendency to be biased by their own guidelines, while D1 and D2 have not used documented guidelines. Furthermore, D4's experience and the problems encountered in developing the "Style Guide", for use by designers within the company and by those who are involved in research over the company's products, partially explain the critical demands made on clarity, explicitness and specificity. This will be discussed in the next section.

11.4.4 Suggestions on guidelines refinement

This section presents a few, select guidelines that require modification. The purpose is to illustrate the sorts of comments made and the changes that were suggested. Three of the guidelines presented here (Guidelines 1, 16 and 21) were described in Section 11.2.2. Appendix 36 presents a summary of the suggestions on individual guidelines that did not meet the clarity and usability criteria.

Space precludes giving details. Thus, for the following guidelines only comments that suggest needed changes will be presented. This means comments that relate to the positive aspects of the guideline or discussions on issues raised in a particular guideline will not be reported. Since a guideline consists of four parts: title, statement (ie. recommendation), example and comment (see Appendix 34), the modifications could relate to any of these. However, only the title and statement will be presented below. The source of the comment will be indicated in brackets.

1 Clearly defined functions for input devices

The functions of each input device should be well rationalised and clearly distinct from each other.

- Qualify the example, that is, speech recognisers have a <u>potential</u> to be best suited (...). [D3]
- Modify the statement: "the type of input should be well rationalised", <u>not</u> the functions of each input device so as to achieve the degree of flexibility that is needed. [D4]
- Indicate in the comment that there are conflicting guidelines. Give explicit examples of that conflict and how to resolve them. [D1, D6]
- Distinguish this guideline clearly from Guidelines 2 and 3. [D7]
- Explain further why certain input devices are suited to the functions mentioned. [D2]

2 Flexible assignment of input devices to data type

Allocation of input mode to data should be flexible so that it does not load user memory.

- Define the terms *user, input mode* and *flexibility*. [D1]
- Qualify the example, that is, "do not split commands between input modes". [D1, D2, D3]
- Explain in the comment that a flexible interface is one whereby all the input devices can do everything. [D2, D4]
- State the criteria for allocating input modes to data type. [D5]

• Specify what goes where so that designers do not have the options to make a choice. [D4]

16 Minimal use of manual data entry

Data entry via manual input mode should be kept to a minimum so that a user can stay with one manual method of entry, and not have to shift to another.

• Define the term *costs* and how they are measured. [D3]

• Clarify what is meant by *minimal entry* and *modality*. [D1]

• Provide a higher level, overriding design principle which would resolve the conflict between minimal entry and shifting of hand between input devices, given that both are in the same statement. [D6]

19 Responsiveness to transducer pen-down

The tablet should remain unresponsive to slight pressure from the transducer as would occur if the user swept the stylus or puck lightly over the tablet surface.

• Change the title as it is not clear. [D4]

• Define what constitutes 'slight pressure'? [D3,D5]

• Qualify the comment by suggesting to designers to try out with users and if they failed to register a response *n* times, use this as a criterion. [D6]

• Explain that this problem is peculiar to a stylus rather than a puck. Also, a puck is more stable than a mouse and stylus in menu/entity selection. [D2]

21 Flexible allocation of system information to displays

Allocate system information flexibly between displays in a dual-screen configuration so that users have the choice of which display to view for the information.

• Clarify the <u>users</u> in this guideline as they could be end users, system designers, etc. [D1]

• Make this guideline generic to single screen systems as well. For example, 'dual-screen' may be seen as equivalent to 'windows' on a single screen. [D4, D1]

• Emphasise the importance of providing alternatives for different user levels. [D4, D7]

• Qualify the comment of the guideline as it is not clear. [D1]

• Suggest the use of screen menus that can be moved about in different parts of the display(s) as desired to increase flexibility. [D2]

• Emphasise that the command line on the primary screen should be more than one line than that currently available in order to provide adequate feedback. [D2]

24 Error messages for verbal repeats

Display a non-disruptive error message if a user repeats an entry that is already recognised.

• Avoid using an asterisk to indicate an error as it may be interpreted as something important. Use other forms of symbol (eg. !) instead. [D3]

• Provide an error message in the same channel as the error occurs. [D7]

• Modify the comment: the non-disruptive message is not an error message since it is not a user error. [D1]

• Indicate that there should be feedback on what the system actually recognises, whether the first entry or the repeat entry. [D2, D3]

• Clarify that the system should accept the repeat but the user must be warned of it and the system must be able to cancel the repeat if it is the same as the first entry. [D2]

• Clarify that this example has not been tested, thus it is a suggestion for further research. [D5]

In summary, several suggestions were made to revise and refine the proposed guidelines. It was suggested that Guidelines 8, 14 and 15 (see Appendix 34) should be treated separately under a section on physical aspects of the system since "they are not concerned with the user interface", although this could not be true given the definition in Section 11.2.1. Also, it was considered necessary to have more guidelines on user guidance of the types mentioned in Guidelines 22 to 24 and on the help facility. The use of a single sentence for each recommendation was seen as useful: it helps to constrain the scope of the guideline. The major modifications relate specifically to:

(1) defining the various terms explicitly given that the guidelines were presented out of their thesis context;
(2) providing more specific examples for implementing the guidelines;
(3) fine tuning of the reasoning behind each guideline;
(4) modifying the titles to reflect the recommendations; and
(5) organising the guidelines into clusters of well-defined and specific topics.

The refinement of the guidelines represents the third stage of the development process. As mentioned earlier (Section 11.1), this will be suggested as further work.

11.4.5 Conclusion

The validation process involving system designers has identified five key issues in the development of guidelines. First, the guidelines need to be clear and explicit in order to be interpreted correctly by their users. Second, the guidelines need to be specific with respect to the examples given, and highly prescriptive in order to reduce unnecessary decisions by system designers. Third, the guidelines need to be supported by empirical evidence in order to ensure that they are implementable. Fourth, the guidelines need to be generic in order to be applied in various domains of application. Fifth, the guidelines need to be delivered in easily accessible forms in order to be highly usable. The proposed guidelines satisfied in part some of the key issues, but there is still scope for improvement.

11.5 SUMMARY

With the aim of generating a set of guidelines that would be usable in configuring and/or designing CAD systems that integrate speech and manual input, twenty-four guidelines were developed using the experimental data and related literature. To determine if these guidelines were usable, 7 system designers were interviewed. On the basis of their interview protocols, each guideline was assessed in terms of its clarity and usability. On the whole, the guidelines were found to be usable but needed revision following suggestions by the designers. This refinement of the guidelines is seen as a third phase in the development cycle.

NEXT CHAPTER HIGHLIGHTS

Chapter 12 concludes the thesis by summarising and discussing the thesis outputs. Gaps in the research will be identified. A review of the research approach will be made in order to assess the accomplishment of the thesis goals. Problems encountered and matters arising from the investigation for future research will also be discussed.

CHAPTER 12

Thesis Outputs and Future Directions: A Summary and Discussion

Overview

12.1 Introduction

- 12.1.1 A review of the thesis in terms of the framework
- 12.1.2 Limitations of the thesis findings

12.2 Thesis outputs: summary and discussion

- 12.2.1 Significant findings of problem-solution analysis
- 12.2.2 Behaviour-based methodology as a research tool
- 12.2.3 Blackboard model of system behaviour as an analytical tool
- 12.2.4 Human factors guidelines as a design tool
- 12.2.5 Conclusion: implications for system design

12.3 Problems encountered

12.4 Future directions in user interface development

- 12.4.1 Dialogue design
- 12.4.2 Development of a behavioural design database
- 12.4.3 Development of ergonomic design standards

Final remarks

CHAPTER 12

Thesis Outputs and Future Directions: A Summary and Discussion

OVERVIEW

The thesis has addressed the question of human factors of integrating speech and manual input in CAD systems. The investigation has produced four major outputs. First, the experimental findings with respect to the documentation of problems and the investigation of solutions. Second, the behaviour-based methodology which outlined an approach towards understanding the problem-solution relationship. Third, the blackboard model of system behaviour which provided an analytical tool for understanding the discrepancies between normative and observed behaviour. Fourth, the human factors guidelines developed for use in configuring or designing multimodal CAD systems. Each output will be discussed in turn.

This chapter will also examine gaps in the research and make suggestions for further research. Future directions in user interface development will be identified as possible solutions to the problems associated with the transfer of technology, and the provision of support facilities for human factors designers. Problems encountered in the investigation, particularly those pertaining to the use of speech input, will also be discussed.

12.1 INTRODUCTION

The goal of the thesis was to investigate human factors of integrating speech and manual input in CAD systems. To accomplish this goal, the thesis employed a research approach that began with initial problem specification in the workplace, then proceeding to the investigation of solutions in laboratory contexts, concluding with an application of the findings to develop guidelines and validating them in the workplace. This approach will be reviewed in the light of Long's (1989) framework for ergonomic activities.

12.1.1 A review of the thesis in terms of the framework

This description of the thesis in terms of Long's (1989) framework is to provide a coherent way of understanding the research activities (including concepts and methods) and their contributions to the thesis development.

The domain in which the research applies is CAD, which in turn is concerned with human-computer interaction (HCI). The role of this HCI research is to provide, in part, a means of understanding the optimisation of system behaviour in system development. System development (or interaction development practice) is the process required to develop specifications of computer and user behaviours, such that, once implemented, a desired performance of work results from their interaction (Long, 1989). The activity of *specification* produces a structural and behavioural representation of the system to support the interaction. The activity of *implementation* instantiates the specification in a particular system which supports the interaction (Long, 1989). Modelling the support for system development requires a framework which characterises the activities inherent in the development practice.

In the framework (see Figure 1.1, Chapter 1), the *real world* is contrasted with a *representational world* whose function is to support intended change in the real world. In terms of the framework, the real world of system development expresses the classes of intentional change of an existing system (ie. unitary manual input CAD systems) into a hybrid system (ie. integrated speech-manual CAD systems). Therefore, the changed interactive system would be the new development of an existing CAD system. Analysis of the existing system produces a set of intermediary representations, reflecting the activities undertaken to support the development practice.

In terms of the framework, the HCI practice generally consists of two main activities. One activity acquires knowledge about CAD systems - users, computers, tasks and environments - which in turn produces an *acquisition representation* in Figure 1.1. In the context of this thesis, knowledge was acquired from: (1) the documentation of the problems relating to the unitary use of manual input devices in CAD systems through observation of CAD experts at work (Chapter 6); and (2) the experimental investigation of human factors solutions to the problems of non-optimal behaviour and performance (Chapters 8 to 10). The problem-solution analysis required a methodology that enabled empirical evaluation of CAD systems (Chapter 4), and a demonstrator CAD system whose performance was optimised (Chapter 7). To relate the solution to the problem, a model of system behaviour was developed (Chapter 5). This model, based on the Blackboard framework of design, was a particularisation of the science support representation. Generalising the experimental findings resulted in a science base knowledge.

The other activity applies the knowledge. In this thesis, the development of human factors guidelines for the configuration or design of multimodal CAD systems is a utilisation (or application) of the experimental findings from the science base (Chapter 11). This activity produces an *application representation* in terms of the framework.

Synthesis of the above application representation with the real world would produce the changed interactive system. However, this thesis investigation did not attempt to apply the guidelines in developing novel CAD systems in the real world. Nevertheless, there is sufficient evidence that the guidelines produced are potential design tools as validated by system designers in the workplace. Given the resource constraints of the thesis, and the need to revise the guidelines via iteration, such synthesis activity can only occur at a later time. This means recommending synthesis for further research.

It should be pointed out that although Long's framework is useful and complete for modelling the research activity, given the available resources, some transformations and representations were more complete than others. For example, the acquisition of knowledge through the problem-solution analysis, using the experimental system as a research tool, was examined in more detail than the application of the proposed guidelines in the workplace. The weaknesses of the approach are discussed in the next section.

12.1.2 Limitations of the thesis findings

The major limitation of the thesis concerns the extent to which the findings are generalisable. In human factors work, the word *generalisation* refers to two different situations. The first is to generalise or extrapolate from a particular study to some specific application. The second is to apply the findings over a wide range of situations (Chapanis, 1988). In terms of this thesis, the first appears to be more relevant than the second, but there are potential problems in both terms. In particular, the guidelines that were derived from the thesis investigations have limited scope, hence they may not be generalisable to application domains other than CAD. Within CAD application, the proposed guidelines may not be generic to systems other than PC-based CAD systems used in the study. The problems stem largely from the methodology used to investigate the effects of CAD systems. The focus on a limited number of independent variables (mainly the manipulation of input devices and data configuration) and the use of a small-scale test system, including the speech recognition device, questions the external validity of the findings.

The research has one theme - the integration of speech and manual input devices as a bimodal system. As discussed in Chapter 2, a number of human factors variables affect the use of input devices, in particular user characteristics (eg. skill level, experience, etc.). Although heterogeneity was 'designed' into the research by using the three user groups - experts, novices and naive users, no attempt was made to compare them in any instance. Allwood (1986), in a literature review on computer novices, reported that novices have less, and more fragmented knowledge, spend less time encoding the task and do so in a way that is more determined by the surface features of the problem or information given, compared with experts. Novices in general make more errors and have greater difficulties finding them than experts. These differences in knowledge and skill might have consequences on the optimisation of behaviour and performance.

In the context of this thesis, the novices not only have limited knowledge in CAD but they were also inexperienced with the speech device. This probably accounts for the greater recruitment of Text-screen knowledge (for feedback on errors) as opposed to Graphics-screen knowledge. Although there were similarities in the pattern of behaviour between the experts in the Observational study (Chapter 6) and the naive users in the Tablet input group (Experiment 1, Chapter 8), there still remains the problem of proficiency (ie. performance) which was not compared.

Another source of external invalidity is the unpredictable behaviour of the speech recogniser. Due to its poor performance, there is great potential for confounding variables to contaminate the main effect. These include the users' inability to remain consistent in the task voice, noises in the acoustic environment and the positioning of the microphone. Furthermore, the use of a microcomputer CAD system, which differs considerably in capability from the more powerful minicomputers and mainframes, essentially limits the application of the findings to only systems of its kind, although the use of PC-based systems within the CAD environment is commonplace (see Chapter 3).

In view of the above limitations, from an applications perspective, the thesis findings are only generalisable to the intended end users (novices and naive users), providing further work is done to validate and/or replicate the research with expert CAD users. The next section reviews the work undertaken.

12.2 THESIS OUTPUTS: SUMMARY AND DISCUSSION

This section summarises the products of the thesis, namely, the experimental findings, the behaviour-based methodology, the system model of behaviour and the guidelines. Each output will be discussed in terms of the present work (ie. what has been done) and gaps in the research which require further work (ie. what needs doing).

12.2.1 Significant findings of problem-solution analysis

Present work

The thesis adopted a dual approach, namely, the documentation of problems and the investigation of solutions. The observational study of CAD experts at work (Chapter 6) identified the problems of non-optimal behaviour in using unitary manual input in existing CAD systems. In particular, the use incurred more frequent and longer periods of eye gaze to the tablet and keyboard to manipulate the input devices, relative to graphics screen gazing. Additionally, the use caused frequent shifting of the hands between input devices. Similar phenomena were demonstrated by Van der Heiden and Grandjean (1984) and Monk (1986).

In terms of the model, the frequency and duration of Eyes and Hand knowledge recruited to manipulate the input devices expressed the behavioural costs incurred. These costs were considered non-optimal. Therefore, a solution to the problem of non-optimal behaviour was to replace the manual entry of commands and numeric data with unitary speech input.

It was shown in Experiment 1 (Chapter 8) that the use of unitary speech input incurred more frequent and longer periods of eye gaze to the text screen and keyboard. In addition, there was frequent shifting of hands between the tablet and keyboard, with the latter used to reinput commands and numerics. Due to poor speech recognition, there were frequent verbal repeats caused particularly by substitution and rejection errors. The problem of speech confusability affected performance in terms of the data required to produce an entity. Also, there was a tendency for speech input users to rate the system low on acceptability.

In terms of the model, the discrepancy between the predicted frequency and duration of Eyes knowledge recruitment to the text screen and keyboard with observed behaviour suggests that the behavioural costs incurred were substantial. Thus, the use of unitary speech input was also non-optimal. A possible solution then was to integrate speech and manual input within a single system.

In Experiment 2 (Chapter 9), the use of integrated systems, in general, reduced the frequency and duration of keyboard gazing, but the time spent gazing at the text screen was still considerable. Furthermore, dividing the data set between input modes incurred forgetting errors on the users' part. That is, they forgot what data was available within a particular input mode. The integrated systems, however, resolved the performance problem. That is, the use reduced the cost of generating a drawing entity: they were more data efficient.

In terms of the model, the behavioural costs due to longer periods of Eyes knowledge recruitment to the text screen and frequent Speech reentry of data caused by forgetting errors were non-optimal. The costs tended to correlate with production time and production efficiency, suggesting that this configuration of the systems is still non-optimal. Also, it has been shown that increasing graphics screen gazing will improve product quality and reduce production costs. Thus, a solution to the problems documented here was to improve the configurability aspects of the system.

Experiment 3 (Chapter 10) investigated alternative configurations of speech-manual CAD systems by introducing system prompts and backup facilities to aid the user. This manipulation led to increased gazing to the graphics screen in both duration and frequency. There was significant reduction in the time spent looking at the text screen. But the time spent in menu search was substantial. This was partly due to the sensitivity of the transducer, used to select the menu items on screen.

In terms of the model, the behavioural problems documented in previous experiments were resolved: the duration of Eyes knowledge recruitment to the graphics screen was significantly improved, leading to more optimal behaviour. Because users were provided with backup aids to speech entry, this reduced the unnecessary recruitment of Speech knowledge. Speech performance too was improved considerably, from low to moderate recognition threshold, as defined within the thesis.

It should be noted that the above findings are a subset of the significant findings which are not summarised in full here. Other findings of interest include the use of the hand to manipulate the input devices (ie. the idleness issue) and its relationship with production costs. Such findings would have specific implications for the configuration of system components, in terms of optimising the use of input devices so as to distribute the load. The general claim that CAD is a hands-eyes busy task (see Chapters 1 to 3) was confirmed in this thesis; removing this component of the task (ie. making the hands less busy) was detrimental to performance.

It should be made very clear that the design or configuration of integrated speech-manual CAD systems must consider two central issues. First, the effect of the configuration on both behaviour <u>and</u> performance aspects of device use, independent of each other as well as the correlation between them. Second, the importance of identifying user needs in order to ensure that the developed system satisfies their requirements. The latter implies that the design of new CAD systems should have end users as the central focus of the design process. Although system design involves several tradeoff issues such as time, effort and money, ignoring some of the concerns raised here will result in a less usable system. A usable system is one that incurs minimal costs to the user in learning and using the system to support task performance. This thesis has demonstrated some of these concerns.

Gaps in the research

Within the constraints of the thesis, it was not possible to investigate some issues in depth. The following suggestions are therefore aimed to provide better understanding of related design issues that may have consequences for design behaviour and task performance.

- combining graphics screen prompts with a screen menu as an alternative speech-plus-manual CAD system. This might increase further graphics screen gazing and lead to more optimal behaviour in a dual-screen configuration.
- using a single screen system but with multiple windows. This should eliminate between-

screen transitions altogether but with some potential costs to space and clarity on the single screen.

- using auditory prompts (ie. computer-generated speech) in place of visual prompts, or a combination of visual and auditory. This should optimise the use of the human sensory channels given the compatibility between I/O modalities (see Chapter 8). Also, this might help to identify the form of prompts that best supports CAD task performance.
- using different forms of screen menu as a backup facility. The combination of speech input and different types of screen menu (eg. pull-down, pop-up) might help to speed up menu search.

12.2.2 Behaviour-based methodology as a research tool

Present work

The thesis employed a methodology which uses behavioural and performance data as its primary source of information for empirical evaluation of CAD systems. The behavioural data was derived from a scoring of behaviour protocols recorded on videotapes. To process and analyse the behaviour data, the methodology suggests the use of a computerised technique for video analysis. This enabled microanalyses of behaviour in terms of frequency and duration metrics. Other aspects of the methodology include the use of modelling techniques to represent system behaviour and the use of statistical tools to test the hypothesised effects of system on the derived measures as well as the relationships between them.

A behaviour-based methodology sufficiently sensitive to the above concerns can therefore potentially serve to delineate the types of system components that are appropriate for the CAD systems either currently available or likely to be become available. Evidence (eg. Card, Moran & Newell, 1983; Sharit & Cuomo, 1988) suggests that the combined use of behaviour and performance measures in CAD evaluation offers better insight into design activities. It has been shown in this thesis that the methodology supported the problem-solution analysis, in particular, the mapping of the relations between the system 'treatment' (ie. independent variable) and the intended outcome (ie. dependent variables). To this end, it could be said that the methodology supports an understanding of design behaviour and task performance which, in turn, is fundamental to optimal system design.

Gaps in the research

The methodology, however, is not complete with respect to the categories of behaviour that could be investigated. For example, it is also important to understand the cognitive aspects of CAD performance (eg. Whitefield & Warren, 1989; Ballay, 1988). The type of cognitive behaviour that is examined in the methodology concerns memory load, that is, forgetting. An evaluation of cognitive processing (ie. decision making, problem solving) might help to explain why users perform a certain action and how they process information from the display. For example, it is not always clear why users actually gaze at a particular target. Their intentions could only be inferred from the use of specific commands and the type of data entry (ie. graphical or alphanumeric data).

The use of the computerised technique in scoring behaviour protocol raises several issues concerning reliability and validity of measurements (see Chapter 4). Although precautions were taken to minimise the effects of extraneous factors from confounding the measurements, the nature of the recording and scoring could not guarantee this. One way of resolving this issue is to use several well-trained raters to score the data (Laws, 1988). By performing inter-rater reliability tests on the scored data, an index of consistency could be derived.

12.2.3 Blackboard model of system behaviour as an analytical tool

Present work

The characterisation of system behaviour was presented in a system model. The framework for the model was derived from the blackboard metaphor (Chapter 5). The model illustrates the different types of behavioural knowledge recruited during CAD performance. The concepts expressed in the model were determined by the system technology, specifically the four major components of system - user, computer, task and environment. The role of the model was basically analytical. This means the model provided a means of understanding how and why different knowledge might be recruited in task performance. Additionally, the model has helped to identify the triggers for particular behaviours, based on the control knowledge algorithms. These indirectly suggested ways of optimising behaviour for enhancing performance further.

However, the model was not used to drive the research and it does not accumulate knowledge across investigations since the basic types of knowledge (ie. system tools) were the same throughout. However, the range of knowledge within a particular tool type (eg. different types of input devices) could be expanded, depending on the nature of the investigation.

As an explanatory tool, this blackboard model of system behaviour supports an understanding of the problem-solution analysis. It could be said that the model is a novel way of capturing system behaviour at the input/output level of the CAD process. It therefore has the potential to describe system behaviour in applications other than CAD (eg. database management).

Gaps in the research

The model expresses behaviour as occurring in time and space. Therefore, there were two levels of solution intervals, namely time on the temporal dimension and movement space on the spatial dimension. Within the thesis constraints, it was not possible to gather data on movement space. This would include measuring the distance of travel of the hand from the tablet to the keyboard, the visual angle subtended by the eye to the line of sight of various targets, etc. Spatial information is important in determining, for example, the arrangement of items in the display, optimal positioning of input devices within easy reach of the user. For this purpose, the blackboard model could be modified by synthesising the temporal and spatial dimensions as one (see Chapter 5). This should help to simplify the model operations.

12.2.4 Human factors guidelines as a design tool

Present work

The thesis generated a set of human factors guidelines, derived from the experimental data (Chapter 11). The proposed guidelines were intended to provide support in the configuration of speech-plus-manual CAD systems by end users and CAD implementors, and in the design of multimodal CAD systems by system developers. To ensure that the guidelines are usable, they were validated by system designers, who had the expertise in the design of user interfaces for CAD and/or communication systems involving speech. The guidelines were intended for a specific application, and are not intended to generalise to other domains.

The guidelines evolved through the process of develop, validate, and ultimately, revise. The validation process identified pertinent problems in the use of the guidelines, particularly their clarity, specificity and organisation. As such, the guidelines needed refinement following recommendations by the designers. In terms of the thesis, the guidelines are a usable product; in terms of system design, the guidelines are not a specification of a system (ie. a design standard), but a design tool which could help the designer to consider some important aspects of the design process.

Gaps in the research

The proposed guidelines may be improved in four significant ways. First, refinement is needed in the structuring and organising of the guidelines. In other words, a framework is required that would underlie the design of the guidelines, including the terminology used. Providing a glossary for the latter should help to define the terms, making the guidelines self-explanatory. Second, the format for expressing the guidelines needed revision. In order to be highly usable, the guidelines must be delivered in a form that is easy to use and implement. The present form, following Smith and Mosier (1986), is acceptable in so far the structure and presentation style are concerned. However, the examples in the proposed guidelines needed to be more prescriptive and directive based on the empirical evidence of the thesis investigations and other guidelines literature (eg. Shneiderman, 1987). Examples that require further research should be stated as such or dropped altogether. Third, the guidelines should be made generic so that the scope of their application could be extended to more general-purpose systems that integrate speech and manual input devices. Lastly, the guidelines should be highly prescriptive and tested so as to offer better guarantees that their application in system design would lead to a more usable product. This refinement of the guidelines is necessary to achieve a usable design tool.

12.2.5 Conclusion: implications for system design

In terms of the goal of the thesis, it could be concluded that this has been accomplished. However, due to various constraints, only a subset of the problem-solution space could be investigated in detail. This analysis has provided some insights into the use of input devices on a unitary or combined basis. Some issues which could not be investigated were recommended for further research. Such an undertaking is crucial to enable better understanding of potential problems in speech-manual integration.

The various thesis outputs have relevance to, and potential for, the design of integrated speech-manual CAD systems. To be most effective, the design of new systems should be based on a coherent set of assumptions about optimal design behaviour. The blackboard model of system behaviour provides a tool for understanding this. The empirical findings, expressed in the form of guidelines, provide a source of information for the tested assumptions. The methodology binds the research activities together.

The guidelines imply that to promote optimal task performance, systems should be designed to be compatible with the normative behaviour of CAD users under routine design activity. Design of I/O should allow flexibility in both the display and manipulation of information to complement individual differences and preferences in work styles and approaches to problem-solving. Users should be supported with easy to use performance aids, whether in the form of help facilities, memory aids, menus, checklists or prompts. System feedback and error correction strategies, the latter in the form of repair and recovery loops, should be provided to reduce confusability and to aid performance further.

12.3 PROBLEMS ENCOUNTERED

The use of developing technologies - speech recognition and computerised video scoring - inevitably throws up some technical difficulties. These become critical when the developers (eg. manufacturers) themselves may not be aware of these problems. To provide some pointers to researchers who may be involved in similar work, a statement of the problems encountered will be presented, followed by way(s) of resolving them. The main problems concerned the maintainability of speech recognition during experimentation and the recording of behaviour protocols.

<u>Problem 1</u>

Given the confusability aspects of speech input, any noise in the acoustic environment can distort recognition. This includes a range of factors such as inadvertent lip-smacking, sighing, telephone ringing, voices in the background, keyboard pressing, etc. Given the type of computer system used, maintaining speech input on one drive (eg. the hard disk) together with other data increases its probability of being confused with similar-sounding words. The problem would have more serious consequences if the words were confused with commands in the operating system (eg. DOS commands - format, break, etc.). To avoid such mishaps,

- maintain the speech program and vocabulary in a separate drive from other data.
- switch off the recogniser during critical aspects of the task that do not require it.
- unplug the microphone from the computer when not in use.

<u>Problem 2</u>

Recording behaviour protocols can be a problem if a whole range of user actions are required in the analysis. The quality of the recorded image is important to ensure fidelity. The representation of the image is crucial to aid automated scoring. Using two cameras to record the scenario was found to be inadequate as some tradeoffs had to be made, such as whether to record the screen display as opposed to the manipulation of input devices. Both considerations were crucial to behaviour analysis. To avoid such decisions during recording and scoring,

- try using more than two cameras, each recording a specific behaviour but mixed via a vision mixer. For example, one camera on eye movements, another on hand movements, a third on screen image, etc.
- define the behaviour types and identify the start-end of each type, prior to scoring. This should help to avoid ambiguity and inconsistency in scoring.

These problems are not exhaustive; they are raised here to illuminate potential difficulties that may be experienced in similar undertakings.

12.4 FUTURE DIRECTIONS IN USER INTERFACE DEVELOPMENT

This section takes a brief look at future research issues in the development of user interfaces for interactive systems. These issues arise from the work that was undertaken here. The purpose is to identify particular areas where further research efforts are needed to support the development of systems. These include the design of adaptive dialogues to meet user requirements, the development of a database on behaviour and performance data, and the development of ergonomic design standards.

12.4.1 Dialogue design

There has been a great deal of research into the structure, content and style of human-computer dialogue. (Within HCI, the term *dialogue* is often used interchangeably with the term *interface*. Here, the term dialogue will be used to mean the process of communication that occurs at the interface.) This extensive research, however, has not produced any unified theory or explanatory framework (Booth, 1989). Nevertheless, general dialogue design guidelines (eg. McMillan & Moran, 1985; Shneiderman, 1987) have emerged.

The use of a command language in human-computer dialogues is common. The need to adapt the dialogue to the requirements and cultural level of users is much emphasised in the literature (eg. Chapanis, 1975; Sinaiko, 1975; Sperandio, 1987). The goal is to support the transfer of technology (see Chapter 3). Development of dialogues is typically based on the English language. There are several linguistic similarities between English and *Bahasa Malaysia* (the national language of Malaysia). First, both languages are phonetically-based. Second, both use Roman characters. Third, both have rigid syntax. One difference is that a majority of the words in the Malaysian vocabulary is not monosyllabic. This characteristic is useful in developing command languages for speech input since the probability of discriminability is increased. To illustrate the similarities between English and Bahasa Malaysia, two examples are given below.

<u>Example 1</u>

English	She	is	a	stu|dent
B. Malaysia (translation)	Ia	a|da|lah	sa|tu	pe|nun|tut
		copula		*noun*
	'Ia (adalah) penuntut'.			

<u>Example 2</u>

English	You	are	shown	the	book	by	him
B. Malaysia (translation)	An|da	a|da|lah	li|hat|kan	i|tu	bu|ku	o|leh	ia
	Subject	*copula*	*indirect verb*		*noun*		
	'Anda dilihatkan buku itu olehnya'						

Given the above similarities and the state of the art in speech dialogue design, there is potential for a one-to-one transfer of the available technology in developing multimodal systems driven by the Malaysian language. This should involve minimal adaptation of the modelled system, which in turn should reduce the time and effort required for system development. It cannot be denied that 'adaptive' systems should help to: (1) speed up the learning of particular applications; and (2) increase an appreciation of the technology. Therefore, major research efforts are needed to explore the possibilities. The significance of such research to the thesis is in the development of a more adaptive speech interface based on locally-orientated user dialogues.

12.4.2 Development of a behavioural design database

Design is a major area of concern for human factors. One way to support system design is the development of a behaviourally orientated design database. To be useful, the database must have two characteristics:

(1) it can be related easily to the design of the primary system; and

(2) its inputs should be in some quantitative form.

Ideally, the material provided should contain behavioural principles of design, and associated with such principles there should be human performance data, expressed in terms of optimal performance. The data are needed to provide empirical support for the principles. In addition to this, there should also be quantitative relationships and tradeoffs between design parameters in behavioural and performance terms. Lastly, there should be procedures for designers to follow in developing their designs. These procedures, based on a model of design behaviour, could be used to guide the designer in the design process.

Given that very little is still known about design dynamics and how these relate to behavioural inputs, the development of this database would be in the right direction. The work reported here has identified some of the system behaviour variables that could be considered for inclusion in such a database, including their quantitative form.

12.4.3 Development of ergonomic design standards

Alternatively, to support successful design appropriate standards are required. These include national standards framed within guidelines agreed through International Standards Organisation (ISO). Standardization provides economic benefit for the production of one particular product or process, but with varying benefits to associated activities (Tom, 1988). Within HCI, ergonomic design standards relate to design specifications of the user-computer interface. Ergonomic research on aspects of human-computer interaction provides a source for deriving these standards. For example, the development of most German ergonomic standards was based on research projects (Dzida, 1989). To this end, research on human factors issues needs to be stepped up.

The thesis has generated some guidelines for the design of multimodal CAD systems. To become design standards, these guidelines would need to be rigorously applied within the domain. Thus, major efforts need to be made in this direction to ensure that the guidelines are implementable. With sufficient documented evidence, the guidelines may be transformed to become design standards for multimodal systems.

In conclusion, the above issues are aimed to promote the design and development of usable interactive systems. Design standards are vital in ensuring universally accepted products; the behaviour database is important as a source for supporting design activity and for deriving the standards; and lastly, the design of adaptive dialogues is crucial in improving learning and communication.

FINAL REMARKS

A major lesson of this thesis is that it is essential to analyse interactive behaviour concerning device use in terms of human factors - user, computer, task and environment. Design of multi-modal CAD systems must be looked at as a whole, not in terms of the isolated components of the system. The superordinate goal of the design of hybrid systems is to optimise design behaviour and task performance. Given that speech input is not the 'natural' solution to improving human-computer interaction, integrating speech and manual input within a single system seems to offer many advantages and a way forward.

The advantages evidenced in this thesis are: (1) an integrated system allows for increased visual attention to the graphics screen during design activity; (2) eye and hand transitions between manual input devices are reduced; (3) the assignment of devices according to task characteristics (eg. commands to speech input, coordinates and numerics to tablet input and text to keyboard input) enables multimodal interaction with the computer; and (4) the manual input mode can serve as a backup facility to the speech mode, and vice versa. In short, many of the behavioural problems documented with using manual or speech input on a unitary basis were resolved through integrating the two input modes. On this note, speech-manual integration can be regarded as both an intermediate and a long-term solution, depending on the available speech recognition technology. With the current technology, speech is suitable only for a subset of the computer tasks, for example, command entry. This solution is thus an intermediate one. With improved technology, speech could be suitable for a wider range of tasks, such as graphical and alphanumeric entries, besides command entry. This long-term solution would result in a more flexible multimodal system. In both cases, however, appropriate integration is a better solution than using the devices on a unitary basis, irrespective of the state of the art of the individual technologies - speech and manual input. This is because integration optimises the human's perceptual-motor behaviours, resulting in enhanced performance. Speech-manual integration, therefore, offers a promising approach towards improving interactive systems.

BIBLIOGRAPHY

ALLPORT, D.A. (1980). Attention and performance. In G. Claxton (Ed.),*Cognitive Psychology: New Directions*. London: Routledge & Kegan Paul.

ABLETT, S. (1988). Back to the blackboard. *Network,* February edition, 58-63.

AKIN, O. (1984). An exploration of the design process. In N. Cross (Ed.), *Developments in Design Methodology*. Chichester: John Wiley & Sons.

ALLWOOD, C.M. (1986). Novices on the computer: a review of the literature. *International Journal of Man-Machine Studies,* **25**, 633-658.

ANDERSON, J.R. (1983).*The architecture of cognition.* Cambridge: Harvard University Press.

ANTIN, J.F. (1988). An empirical comparison of menu selection, command entry and combined modes of computer control. *Behaviour and Information Technology,* 7(2), 173-182.

ASIAN COMPUTER MONTHLY (1986). August edition, 34.

AUTODESK LIMITED (1985). *AutoCAD User Reference Manual.*

BABER, C., and STAMMERS, R.B. (1989). Is it natural to talk to computers? An experiment using the "wizard of oz" technique. In E.D. Megaw (Ed.),*Contemporary Ergonomics 1989*. London: Taylor & Francis.

BAILEY, P. (1985). Speech communication: the problem and some solutions. In A. Monk (Ed.), *Fundamentals of Human-Computer Interaction*. London: Academic Press.

BAILEY, R.W. (1982). *Human Performance Engineering: A Guide for System Designers.* Englewood Cliffs, New Jersey: Prentice-Hall.

BALLAY, J.M. (1988). An experimental view of the design process. In W.B. Rouse and K.R. Boff (Eds.), *System Design: Behavioral persepctives on designers, tools and organizations.* Amsterdam: North-Holland.

BARNARD, P.J., HAMMOND,N.V., MORTON,J., LONG,J.B., and CLARK,I.A. (1981). Consistency and compatibility in human-computer dialogue. *International Journal of Man-Machine Studies*, **15**, 87-134.

BARR, A., and FEIGENBAUM, E.A. (1983). *The Handbook of Artificial Intelligence.* California: Pitman.

BEGG, V. (1984). *Developing Expert CAD Systems.* London: Kogan Page.

BERGER, M.M. (Ed.) (1982). *Videotape Techniques in Psychiatric Training and Treatment,* Revised edition. New York: Brunner-Mazel.

BICKEL, S. (1987). Taking the right tablets. *3D.*

BILLSDON, R. (1987). CAD on PCs. *PC User,* July edition, 214-216.

BJORN-ANDERSEN, N. (1988). Are 'Human factors' Human? *The Computer Journal,* 31(5), 386-390.

BOLT, R.A. (1980). "Put-That-There": Voice and gesture at the graphics interface.*Computer Graphics*, 14(3), 262-270.

BOOTH, P. (1989). *An Introduction to Human-Computer Interaction.* London: Lawrence Erlbaum Associates.

BRIGHT, P. (1988). AutoCAD enters the third dimension. *3D,* 45-49.

BROADBENT, G. (1979). The development of design methods. *Design Methods and Theories,* 13(1), 41-45.

BUCKLEY, P.K., and LONG, J.B. (1985). Effects of system and knowledge variables on a task component of "teleshopping". In P. Johnson and S. Cook (Eds.), *People and Computers: Designing the Interface.* Cambridge: Cambridge University Press.

BUXTON, W.A.S. (1986). There's more to interaction that meets the eye: some issues in manual input. In D.A. Norman and S.W. Draper (Eds.),*User-centered System Design: New Perspectives on Human-Computer Interaction.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

CAD User (1988). How to avoid the pitfalls when choosing a CAD system. September edition, 56-60.

CAE BULLETIN (1986). Bechtel group publications, December edition, **1**(2).

CARD, S.K., ENGLISH, W.K., and BURR, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics*, **21**, 601-613.

CARD, S.K., MORAN, T.P., and NEWELL, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

CARD, S.K., MORAN, T.P., and NEWELL, A. (1980). The keystroke model for user performance time with interactive systems. *Communications of the ACM*, **23**, 396-410.

CARROLL, J.M. (1982). Learning, using, and designing filenames and command paradigms. *Behaviour and Information Technology*, **1**(4), 327-346.

CARROLL, J.M., THOMAS, J.C., and MALHOTRA, A. (1980). Presentation and representation in design problem solving. *American Journal of Psychology*, **93**, 269-284.

CARROLL, J.M., and ROSSON, M.B. (1985). Usability specifications as a tool in iterative development. In R. Hartson (Ed.), *Advances in Human-Computer Interaction*. Norwood, New Jersey: Ablex.

CHAMBERS TWENTIETH CENTURY DICTIONARY (1985). W & R Chambers Ltd.

CHAPANIS, A. (1965). On the allocation of functions between men and machines. *Occupational Psychology*, **39**, 1-11.

CHAPANIS, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics*,**10**(5), 557-577.

CHAPANIS, A. (1975). Ethnic Variables in Human Factors Engineering. Baltimore: The Johns Hopkins University Press.

CHAPANIS, A. (1988). Some generalizations about generalization. *Human Factors*, **30**(3), 253-267.

CHRISTENSEN, J.M. (1988). Human factors definitions. *Human Factors Society Bulletin,* 31(3), 8-9.

CHRISTIE, B. (1985). *Human Factors of the User-System Interface. A Report on an ESPRIT Preparatory Study.* Amsterdam: Elsevier Science Publishers.

CLANCEY, W.J. (1983). The advantages of abstract control knowledge in expert system design. *Technical Report HPP-83-17.* Stanford University, Stanford, California.

CLARK, P.E.J. (1986). Implementaton of voice input to computer applications. *Journal of Information Technology,* 1(4), 39-56.

CLARKE, A., and ELLGRING, H. (1983). Computer-aided video. In P.W. Dowrick and S.J. Biggs (Eds.), *Using Video.* Chichester: John Wiley.

CLEGG, C.W. (1986). Trip to Japan: Human resources and synergy. *Personnel Management,* August edition, 35-39.

CLEGG, C.W, RAVDEN, S., CORBETT, M., and JOHNSON, G. (1989). Allocating functions in computer integrated manufacturing: a review and a new method. *Behaviour and Information Technology,* 8(3), 175-190.

COATS, R.B., and VLAEMINKE, I. (1987). *Man-Computer Interfaces: An Introduction to Software Design and Implementation.* Oxford: Blackwell Scientific.

COLE, I., LANSDALE, M. and CHRISTIE, B. (1985). Dialogue Design Guidelines. In B. Christie (Ed.), *Human Factors of Information Technology in the Office.* Chichester: JohnWiley & Sons.

COMERFORD, R. (1984). Pointing-device innovations enhance user/machine interfaces. *EDN: Everything Designers Need,* July 26, 54-66.

COMPUTING (1986). September 25, 30-32.

CRAIG, I.D. (1988). Blackboard systems. *Artificial Intelligence Review,* 2, 103-118.

CROSS, N. (Ed.) (1984). *Developments in Design Methodology.* Chichester: John Wiley & Sons.

DAMOS, D.L. (1985). The effect of asymmetric transfer and speech technology on dual-task performance. *Human Factors*, 27(4), 409-421.

DAMPER, R.I. (1984). Voice input aids for the physically disabled. *International Journal of Man-Machine Studies*, **21**, 541-553.

DAMPER, R.I., LAMBOURNE, A.D., and GUY, D.P. (1984). Speech input as an adjunct to keyboard entry in television subtitling. In B. Shackel (Ed.), *Proceedings INTERACT '84*. Amsterdam: North-Holland.

DAVIS, B. (1987). Future Visions. CADCAM in the 1990s. *CADCAM International*, June, 45-50.

DAVIS, R. (1980). Content reference: reasoning about rules. *Artificial Intelligence*, **15**, 223-239.

DAVIS, R., and BUCHANAN, B.G. (1977). Meta-level knowledge: overview and applications, *Proceedings of the 5th IJCAI*, 920-927.

DAVIS, R., and LENAT, D.B. (1982). *Knowledge-based Systems in Artificial Intelligence*. London: McGraw-Hill.

DAVIS, E.G., and SWEZEY, R.W. (1983). Human factors guidelines in computer graphics: a case study. *International Journal of Man-Machine Studies*, **18**, 113-133.

DEGREENE, K.B. (1970). *Systems Psychology*. New York: McGraw-Hill.

DEGREENE, K.B. (1977). Has human factors come of age? *Proceedings of the Human Factors Society*, Santa Monica, CA: Human Factors Society, 457-461.

DOWELL, J. (1986). An investigation of the influence of CAD graphical representation on design performance. Unpublished M.Sc. Thesis, University of London, England.

DOWELL, J., and LONG, J.B. (1988). Human/computer interaction engineering (HCIE). In N. Heaton and M. Sinclair (Eds.), *Designing End-User Interfaces*. Oxford: Pergamon Infotech.

DOWRICK, P.W., and BIGGS, S.J. (1983). *Using Video: Psychological and Social Applications*. Chichester: John Wiley.

DUMAIS, S.T., and LANDAUER, T.K. (1981). Psychological investigations of natural terminology for command and query languages. In A. Badre and B. Shneiderman (Eds.), *Directions in Human/Computer Interaction*. Norwood, NJ: Ablex.

DZIDA, W. (1989). The development of ergonomic standards. *SIGCHI Bulletin*, **20** (3), January, 35-43.

DURETT, J., and STIMMEL, T. (1982). A production-system model of human-computer interaction. In *Proceedings Human Factors in Computer Systems*. Washington: Association for Computing Machinery.

EASON, K. (1988). *Information Technology and Organisational Change*. London: Taylor & Francis.

ECKERSLEY, M. (1988). The form of design processes: a protocol analysis study. *Design Studies*, **9** (2), April, 86-94.

EDWARDS, J.R. (1989). Computer aided manufacturing and worker well-being: a review of research. *Behaviour and Information Technology*, **8** (3), 157-174.

EHRICH, R.W., and WILLIGES, R.C. (1986). *Human-Computer Dialogue Design*. Amsterdam: Elsevier Science.

ERMAN, L.D., and LESSER, V.R. (1980). The Hearsay-II speech understanding system: a tutorial. In W.A. Lea (Ed.), *Trends in Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall.

FAROOQ, M.U., and DOMINICK, W.D. (1988). A survey of formal tools and models for developing user interfaces. *International Journal of Man-Machine Studies*, **29**, 479-496.

FINKELSTEIN, D. (1988). Integrating industry. *Network*, February edition, 8-13.

FOLEY, J.D., WALLACE, V.L., and CHAN, P. (1984). The human factors of computer graphics interaction techniques. *IEEE Computer Graphics and Applications*, **4**(11), 13-48.

FRANKISH, C.R., JONES, D.M., MADDEN, C., WAIGHT, K., and STODDART, J. (1987). Parcel sorting by speech recognition: human factors issues. In D. Diaper and R. Winder (Eds.), *People and Computers III*, Proceedings of the Third Conference of the British Computer Society. Cambridge: Cambridge University Press.

FRANKS, I.M., and NAGELKERKE, P. (1988). The use of computer interactive video in sport analysis. *Ergonomics*, **31**(11), 1593-1603.

GOODWIN, C. (1986). CAD/CAM user disquiet at lack of standards. *Computing*, September 25, 30-32.

GOULD, J.D., and ALFARO, L. (1984). Revising documents with text editors, handwriting-recognition systems, and speech-recognition systems. *Human Factors*, **26**(4), 391-406.

GOULD, J.D., CONTI, J., and HOVANYECZ, T. (1983). Composing letters with a simulated listening typewriter. *Communications of the ACM*, **26**(4), 295-308.

GREEN, T.R.G., PAYNE, S.J., MORRISON, D.L., and SHAW, A. (1983). Friendly interfacing to simple speech recognizers. *Behaviour and Information Technology*, **2**(1), 23-38.

GREENSTEIN, J.S., and ARNAUT, L.Y. (1986). Human factors aspects of manual computer input devices. In G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics*. New York: John Wiley & Sons.

GREENSTEIN, J.S., and LAM, S.T. (1985). An experimental study of dialogue-based communication for dynamic human-computer task allocation. *International Journal of Man-Machine Studies*, **23**, 605-621.

GREEN,T. R. G., and CLARK, G. (1981). 'Checkout: heuristics speechlink'. *Personal Computer World* , June edition.

GROOVER, M.P., and ZIMMERS, E.W. (1984). *CAD/CAM: Computer-Aided Design and Manufacturing*. Englewood Cliffs, NJ: Prentice-Hall.

HAASE, B. (1989). Living in the nightmarish world of the 3rd dimension. *PC Week*, **5**(27), July 18, 16-17.

HALL, A.D., and FAGEN, R.E. (1956). Definition of system. In L. von Bertalanffy and A. Rapoport (Eds.), *General Systems Yearbook of the Society for the Advancement of General Systems Theory*, **1**, 18-28.

HALLER, R., MUTSCHLER, H., and VOSS, M. (1985). Comparison of input devices for correction of typing errors in office systems. In B.Shackel (ed.), *Human-Computer*

Interaction - INTERACT '84. Amsterdam: Elsevier Science Publishers.

HAMMOND, N., JORGENSEN, A., MACLEAN, A., BARNARD, P., and LONG, J.B. (1983). Design practice and interface usability: evidence from interviews with designers. In A. Janda (Ed.), *Human Factors in Computing Systems*, CHI'83 Proceedings, Boston: ACM, 40-44.

HAPESHI, K., and JONES, D.M. (1988). The ergonomics of automatic speech recognition interfaces. In D.J. Oborne (Ed.), *International Reviews of Ergonomics*. London: Taylor & Francis.

HARRISON, J.A., HOBBS, G.R., HOWES, J.R., COPE, N., and WRIGHT, G.R. (1986). Machine supported voice dialogue used in training air traffic controllers. *IEEE*, 110-115.

HAUPTMANN, A.G., and RUDNICKY, A.I. (1988). Talking to computers: an empirical investigation. *International Journal of Man-Machine Studies*, **28**, 583-604.

HAYES-ROTH, B. (1983). The blackboard architecture: a general framework for problem solving? HPP report no. HPP-83-30, Stanford University, Computer Science department.

HAYES-ROTH, B. (1985). A blackboard architecture for control. *Artificial Intelligence*, **26**, 251-321.

HAYES-ROTH, B., and HAYES-ROTH, F. (1979). A cognitive model of planning.*Cognitive Science*, **3**, 275-310.

HAYES-ROTH, F., and LESSER, V.R. (1977). Focus of attention in the Hearsay-II speech understanding system. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, MA.

HEAP, N. (1986). Manufacturing a real growth area. *PC User*, February edition, 156-158.

HUBKA, V. (1983). Design tactics - Methods and working principles for design engineers. *Design Studies*, **4**(3), 188-195.

JOHN. P.A. (1988). The ergonomics of computer aided design within advanced manufacturing technology. *Applied Ergonomics*, **19**(1), 40-48.

JOHNSON, P., LONG, J., and VISICK, D. (1985). Voice versus keyboard: use of a comparative analysis of learning to identify skill requirements of input devices. In M.D. Harrison and A.F. Monk (Eds.), *People and Computers: Designing for Usability*. Cambridge: Cambridge University Press.

JOHNSTON, J.M., and PENNYPACKER, H.S. (1980). *Strategies and Tactics of Human Behavioral Research* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

JONES, J.C. (1970). *Design Methods: seeds of human futures*. Chichester: Wiley.

KANTOWITZ, B.H., and SORKIN, R.D. (1987). Allocation of functions. In G. Salvendy (Ed.), *Handbook of Human Factors*, New York: John Wiley & Sons, 355-369.

KARAT, J., McDONALD, J.E., and ANDERSON, M. (1986). A comparison of menu selection techniques: touch panel, mouse and keyboard. *International Journal of Man-Machine Studies*, 25, 73-88.

KASS, R., and FININ, T. (1988). A general user modelling facility. In *Proceedings CHI 88*, 145-150.

KATO, Y., and TSURUTA, S. (1981). Voice input for CAD systems. In T. Sata and E. Warman (Eds.), *Man-Machine Communication in CAD/CAM*. Amsterdam: North-Holland.

KELSO, J.A.S. (1982). *Human Motor Behavior: An Introduction*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

KELWAY, P. S. (1986). Architecturally speaking. The use of speech technology in architectural design. In A. Smith (Ed.), *Knowledge Engineering and Computer Modelling in CAD*, Proceedings of CAD 86. London: Butterworths.

KERR, B. (1983). Memory, action and motor control. In R.A. Magill (Ed.), *Memory and Control of Action*. Amsterdam: North-Holland.

KIERAS, D., and POLSON, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, **22**, 365-394.

KROEMER, K.H.E. (1972). Human Engineering the keyboard. *Human Factors*, **14**(1), 51-63.

LANG, T. (1985). Design and the personal computer. *PC User*, April edition, 106-118.

LAWS, J.V., SUMMERFIELD, A.B., WATSON, J.P., and ELTON, S. (1986). VITAS: Video Interactive Techniques for Automated Scoring. A frame-accurate microcomputer-controlled video system. Unpublished paper, Department of Psychology, Birkbeck College.

LAWS, J.V. (1988). Computerisation of video analysis techniques: implications for information technology, support and training. Paper presented at the International Conference on "Computerisation for Development - The Research Challenge", 16-18 August, Kuala Lumpur, Malaysia.

LEA, W.A. (1980). *Trends in Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall.

LEA, W.A. (1982). Problems in predicting performances of speech recognisers. In S.D. Harris (Ed.), *Proceedings of the Workshop on Standardization of Speech I/O Technology*. National Bureau of Methods.

LEESLEY, M.E. (1978). The uneven acceptance of CAD. In *Proceedings of CAD 78*, Third International Conference and Exhibition on Computers in Engineering and Building Design. Guilford: IPC Science and Technology Press.

LEGGETT, J., and WILLIAMS, G. (1984). An empirical investigation of voice as an input modality for computer programming. *International Journal of Man-Machine Studies*, **21**, 493-520.

LENAT, D.B., DAVIS, R., DOYLE, J., GENESERETH, M., GOLDSTEIN, I., and SCHROBE, H. (1983). Reasoning about reasoning. In F. Hayes-Roth, D.A, Waterman and D.B. Lenat (Eds.), *Building Expert Systems*. Reading, MA: Addison-Wesley.

LERA, S. (1983). Synopses of some recent published studies of the design process and designer behaviour. *Design Studies*, **4**, 133-140.

LESSER, V.R., and CORKILL, D. (1981). Functionally accurate cooperative distributed systems.*Cybernetics*, 1, 81-96.

LONG, J.B. (1987a). Cognitive ergonomics and human-computer interaction. In P. Warr (Ed.), *Psychology at Work*. Harmondsworth: Penguin.

LONG, J.B. (1987b). A framework for user models. In E. Megaw (Ed.), *Contemporary Ergonomics 1987*. London: Taylor & Francis.

LONG, J.B. (1987c). The application and use of a framework for paradigm models. In E. Megaw (Ed.), *Contemporary Ergonomics 1987*. London: Taylor & Francis.

LONG, J.B. (1987d). Computers and People - Matching Machines to Minds. Inaugural Lecture, University College London, England, 19 March.

LONG, J.B. (1989). Cognitive ergonomics and human-computer interaction: an introduction. In J. Long and A. Whitefield (Eds.), *Cognitive Ergonomics and Human-Computer Interaction*. Cambridge: Cambridge University Press.

LONG, J. B., and JOHNSON, P. (1982). The evaluation of data input devices for parcel and packet sorting. Phase I: optimisation of the devices, Report to the Post Office Engineers.

LONG, J. B., NIMMO-SMITH, I., and WHITEFIELD, A. D., (1983). Skilled typing - a characterisation based on the distribution of times between responses. In W.E. Cooper (Ed.), *Cognitive Aspects of Skilled Typewriting*. New York: Springer-Verlag.

LONG, J.B., WHITEFIELD, A. D., and DENNETT, J. (1984). The effect of display format on the direct entry of numerical information by pointing. *Human Factors*, 26(1), 3-17.

LYNCH, S. (1984). Glass quality inspection system. *Speech Tech '85*, 154-156.

MACKAY, W.E., GUINDON, R., MANTEL, M.M., SUCHMAN, L., and TATAR, D.G. (1988). Video: Data for studying human-computer interaction. *CHI'88 Conference Proceedings, Human Factors in Computing Systems*, Washington: ACM Press, 133-137.

MAGUIRE, M. (1988). Developments in human/computer interaction hardware. In N. Heaton and M. Sinclair (Eds.), *Designing End-User Interfaces - State of the Art Report*. Exeter:

Pergamon Infotech.

MAJCHRZAK, A., COLLINS, P., and MANDEVILLE, D. (1986). A quantitative assessment of changes in work activities resulting from computer-assisted design. *Behaviour and Information Technology*, 5(3), 259-271.

MAJCHRZAK, A., CHANG, T.C., BARFIELD, W., EBERTS, R., and SALVENDY, G. (1987). *Human Aspects of Computer-Aided Design*. Philadelphia: Taylor & Francis.

MALAYSIAN BUSINESS (1986). Upward integration way to go, May 1, 36-38.

MARTIN, G.L. (1989). The utility of speech input in user-computer interfaces. *International Journal of Man-Machine Studies*, **30**, 355-375.

MARTIN, T.B. (1976). Practical applications of voice input to machines. *Proceedings of the IEEE*, 64(4), 487-501.

MARTIN, T.B., and WELCH, J.R. (1980). Practical speech recognisers and some performance effectiveness parameters. In W.A. Lea (Ed.), *Trends in Speech Recognition*. New Jersey: Prentice-Hall, 24-38.

MASSEY, J. (1988). Dimensions on design. *Personal Computer Magazine*, 1(8), September edition, 88-94.

McCAULEY, M.E. (1984). Human factors in voice technology. In F.A. Muckler (Ed.), *Human Factors Review: 1984*. Santa Monica, California: The Human Factors Society, Inc.

McKENZIE, J. (1988). Guidelines and principles of interface design. In N. Heaton and M. Sinclair (Eds.), *Designing End-User Interfaces*. Oxford: Pergamon Infotech.

McLEOD, R.W. (1988). Human factors issues in speech recognition. In E.D. Megaw (Ed.), *Contemporary Ergonomics 1988*. London: Taylor & Francis.

McMILLAN, T.C., and MORAN, B.P. (1985). Command line structure and dynamic processing of abbreviations in dialogue management. *Interfaces in Computing*, 3, 249-257.

MEISTER, D. (1971). *Human Factors: Theory and Practice*. New York: Wiley-Interscience.

MEISTER, D. (1985). The two worlds of human factors. In R.E. Eberts and C.G. Eberts (Eds.), *Trends in Ergonomics/Human Factors II*. Amsterdam: Elsevier Science Publishers.

MEISTER, D. (1987). Systems design, development and testing. In G. Salvendy (Ed.), *Handbook of Human Factors*, New York: John Wiley & Son, 17-42.

MEISTER, D. (1988). A cognitive theory of design and requirements for a behavioral design aid. In W.B. Rouse and K.R. Boff (Eds.), *System Design*. New York: North-Holland.

MICROPHONICS TECHNOLOGY CORPORATION (1984). *Pronounce User Manual.*

MILNER, N. (1988). Input devices and WIMP interfaces. In T. Rubin (Ed.), *User Interface Design for Computer Systems*. Chichester: Ellis Horwood Ltd.

MOHD KHALID, H. (1981). Central Task Demand and Peripheral Visual Performance. Unpublished M.Sc Thesis, Monash University, Melbourne, Australia.

MOHD KHALID, H. (1985). Effects of central task demand on peripheral visual performance. In D. Oborne (Ed.), *Contemporary Ergonomics 1985*. London: Taylor & Francis.

MOHD KHALID, H. (1988). Integrating speech and manual input - towards a hybrid solution: the case of computer aided design. Paper presented at the International Conference on "Computerisation for Development - The Research Challenge", 16-18 August, Kuala Lumpur, Malaysia.

MONK, A. (1986). Mode errors: a user-centred analysis and some preventative measures using keying-contingent sound. *International Journal of Man-Machine Studies*, **24**, 313-327.

MORAN, T.P. (1981). The Command Language Grammar: a representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies*, **15**, 3-50.

MORRISON,D.L., GREEN, T.R.G., SHAW, A.C., and PAYNE, S.J. (1984). Speech-controlled text-editing: effects of input modality and of command structure. *International Journal of Man-Machine Studies*, **21**, 49-63.

MORTON, J., BARNARD, P.J., HAMMOND, N.V., and LONG, J. (1979). Interacting with the computer: a framework. In E.J. Boutmy and A. Danthine (Eds.),*Teleinformatics '79*.

Amsterdam: North-Holland.

MURRAY, D.M. (1988). Report on user models, deliverable D1. Alvey Report 110, National Physical Laboratory, United Kingdom.

NADLER, G. (1989). Design processes and their results. *Design Studies*, **10** (2), April, 124-127.

NEESHAM, C. (1987). Digitizers Digested. *CADCAM International*, June edition, 53-55.

NEWELL, A. (1968). On the analysis of human problem solving protocols. In J.C. Gardin and B. Jaulin (Eds.), *Calcul et Formalization dans les Sciences de L'Homme*, Paris.

NEWELL, A. (1969). Heuristic programming: ill-structured problems in progress. *Operations Research*, **3**.

NEWELL, A., and SIMON, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, New Jersey: Prentice-Hall.

NEWELL, A.F. (1985). Speech - the natural modality for man-machine interaction? In B. Shackel (Ed.), *Human-Computer Interaction - INTERACT '84*. Amsterdam: Elsevier Science Publishers.

NEWMAN, W.M., and SPROULL, R.F. (1979). *Principles of Interactive Computer Graphics*. New York: McGraw-Hill.

NICKERSON, R.S. (1986).*Using Computers. Human Factors in Computer Systems*. Cambridge, Massachusetts: The MIT Press.

NIESEWAND, N. (1986). *Bedrooms and Bathrooms*. London: Conran Octopus Limited.

NORMAN, D.A. (1983). Design rules based on analysis of human error.*Communications of the ACM*, **26**, 254-258.

NORMAN, D.A. (1986). Some observations on mental models. In D.A. Norman and S.W. Draper (Eds.), *User Centred System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, New Jersey: Lawrence Erlbaum.

NORMAN, D.A., and SHALLICE, T. (1985). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schwartz and D. Shapiro (Eds.), *Consciousness and Self-regulation: Advances in Research.Volume 4*. New York: Plenum.

NOYES, J.M., and FRANKISH, C.R. (1987). Review of automatic speech recognition applications. Report for Alvey Contract MMI 103 - *Human Factors in the Design of Speech Systems Interfaces.*

NYE, J.M. (1982). Human factors analysis of speech recognition systems. *Speech Technology,* April, 50-57.

OLINS, W. (1985). Management by design. *Management Today,* February edition, 62-69.

OPEN SYSTEMS GROUP (1981). *Systems Behaviour*. London: Harper & Row and The Open University Press.

OWEN, W. (1988). Where CAD really pays. *3D,* October edition, 22-24.

PEARSON, M. (1988). *Network,* May edition, 66.

PHEASANT, S. (1988). *Bodyspace: Anthropometry, Ergonomics and Design*. London: Taylor & Francis.

PIKAAR, R.N. (1989). Situation analysis of design tasks for CAD systems. *Behaviour and Information Technology,* 8(3), 191-206.

PIPES, A. (1986). When will the Din CAD mean design? *Computer Weekly,* November 13, 50-51.

POLLER, M., and GARTER, S. (1984). The effect of modes on text editing by experienced editor users. *Human Factors,* 26(4), 449-462.

POOCK, G.K. (1982). Voice recognition boosts command terminal throughput. *Speech Technology,* 1(2), 36-39.

POOCK, G.K., SCHWALM, N.D., and ROLAND, E.F. (1982). Use of voice recognition equipment with stenographer masks. Monterey, CA: Naval Postgraduate School.

POST, E. (1943). Formal reductions of the general combinatorial problem. *American Journal of Mathematics*, **65**, 197-268.

RAKER, D., and RICE, H. (1985). *Inside AutoCAD*. Thousand Oaks, California: New Riders Publishing.

RICH, E. (1983). Users are individuals: individualizing user models. *International Journal of Man-Machine Studies*, **18**, 199-214.

RICKETT, J.D. (1987). An investigation and evaluation into the "usability" of human-computer interfaces using a typical CAD system. Unpublished PhD thesis, Council for National Academic Awards, United Kingdom.

RITCHIE, G.J., and TURNER, J.A. (1975). Input devices for interactive graphics. *International Journal of Man-Machine Studies*, **7**, 639-660.

RITTEL, H.W.J. (1984). Second generation design methods. In N. Cross (Ed.), *Developments in Design Methodology*. Chichester: John Wiley & Sons.

RIVERS, R. (1989). Embedded user models - where next? Paper presented at the Conference on "Simulation in the development of user interfaces", 18-19 May 1989, Brighton, England.

ROHR, G., and TAUBER, M.J. (1984). Representational frameworks and models for human-computer interfaces. In G.C. Van der Veer, M.J. Tauber, T.R.G. Green and P. Gorny (Eds.), *Readings on Cognitive Ergonomics - Mind and Computers*, Proceedings of the 2nd European Conference. Berlin: Springer-Verlag.

ROONEY, J., and STEADMAN, P. (1987). *Principles of Computer-Aided Design*. Pitman with the Open University.

ROSE, C. (1988). The design exchange. *Network*, May edition, 66-73.

ROSENMAN, M.A., GERO, J.S., HUTCHINSON, P.J., and OXMAN, R. (1987). Expert systems applications in computer aided design. Proceedings of International Conference on Expert Systems.

ROTH, D. (1988). Is the master serving the slave? *3D*.

ROUSE, W.B., and BOFF,K.R. (1988). Designers, tools, and environments: state of knowledge, unresolved issues, and potential directions. In W.B. Rouse and K.R. Boff (Eds.), *System Design: Behavioral persepctives on designers, tools and organizations*. New York: North-Holland.

RUBIN, T. (1988). *User Interface Design for Computer Systems*. Chichester: Ellis Horwood Ltd.

RUMELHART, D.E. (1976). Toward an interactive model of reading. Technical Report 56, Center for Human Information Processing, University of California at San Diego, California.

RZEVSKI, G. (1984). Impact of information technology on design methods. In R. Langdon and G. Mallen (Eds.), *Design Policy: Volume 6 - Design and Information Technology*. London: Design Council.

SCAPIN, D.L. (1981). Computer commands in restricted natural language: some aspects of memory and experience. *Human Factors*, 23, 365-375.

SCHAFFITZEL, W., and KERSTEN, U. (1985). Introducing CAD systems. Problems and the role of user-developer communication in their solution. *Behaviour and Information Technology*, 4(1), 47-61.

SCHMANDT, C. (1985). Voice communication with computers. In R. Hartson (Ed.), *Advances in Human-Computer Interaction*. Norwood, New Jersey: Ablex Publishing Co.

SCHMANDT, C., and HULTEEN, E.A. (1982). The intelligent voice interactive interface. *Proceedings, Human Factors in Computer Systems*. ACM National Bureau of Standards, 363-366..

SCHMIDT, R.A. (1982). *Motor Control and Learning: a Behavioral Emphasis*. Champaign, Illinois: Human Kinetics Publishers.

SCHOENFELD, A.H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science*, 7, 329-363.

SENKER, P., and ARNOLD, E. (1984). Learning to use CAD. In R.Langdon and G. Mallen (Eds.),*Design Policy: Volume 6 - Design and Information Technology*. London: Design Council.

SHARIT, J., and CUOMO, D.L. (1988). A cognitively based methodology for evaluating human performance in the computer-aided design task domain. *Behaviour and Information Technology*, 7(4), 373-397.

SHERIDAN, T.B. (1988). Please, no more "optimum MMI's"; Or, please watch your language. *Human Factors Society Bulletin*, 31(6), 4.

SHNEIDERMAN, B. (1980). *Software Psychology: Human factors in computer information systems*. Cambridge, MA: Winthrop.

SHNEIDERMAN, B. (1982). Systems message design: Guidelines and experimental results. In A. Badre and B. Shneiderman (Eds.), *Directions in Human/Computer Interaction*. Norwood, NJ: Ablex.

SHNEIDERMAN, B. (1987). *Designing The User Interface: Strategies for effective human-computer interaction*. Reading, Massachusetts: Addison-Wesley.

SHUTOH, T., TSURUTA, S., KAWAI, R., and SHUTOH, M. (1984). Voice operation in CAD system. In H.W. Hendrick and O. Brown, Jr. (Eds.), *Human Factors in Organisational Design and Management*. Amsterdam: Elsevier Science Publishers.

SIMONS, G.L. (1982). *Computers in Engineering and Manufacture*. Manchester: National Computing Centre Publications.

SIMPSON, C.A., McCAULEY, M.E., ROLAND, E.F., RUTH, J.C., and WILLIGES, B.H. (1985). System design for speech recognition and generation. *Human Factors*, 27(2), 115-141.

SINAIKO, H.W. (1975). Verbal factors in human engineering: some cultural and psychological data. In A. Chapanis (Ed.), *Ethnic Variables in Human Factors Engineering*. Baltimore: The Johns Hopkins University Press.

SINGER, R.S. (1980). *Motor Learning and Human Performance: an Application to Motor Skills and Movement Behaviors*. New York: MacMillan.

SINGLETON, W.T. (1974). *Man-Machine System*. Harmondsworth: Penguin.

SLATER, P.J.B. (1978). Data collection. In P.W. Colgan (Ed.), *Quantitative ethology*. New York: Wiley.

SMITH, H.T. (1980). Human-computer communication. In H.T. Smith and T.R.G. Green (Eds.), *Human-computer Interaction*. London: Academic Press.

SMITH, S.L., and MOSIER, J.N. (1986). *Guidelines for Designing User Interface Software*. Bedford, Massachusetts: Mitre Corporation.

SPENCE, R. (1976). Human factors in interactive graphics. *Computer Aided Design*, **8**(1), 49-53.

SPERANDIO, J.C. (1987). Software ergonomics of interface design. *Behaviour and Information Technology*, **6**(3), 271-278.

SPSS INC. (1988). *SPSS/PC+ V2.0 Base Manual.*

STARR, A.F. (1987). Speech recognition technology: a review. Report for Alvey Contract MMI 103 - *Human Factors in the Design of Speech Systems Interfaces.*

SUMMAGRAPHICS CORPORATION (1984). *Summasketch User's Handbook.*

SUMMERFIELD, A.B. (1983). Recording social interaction. In P.W. Dowrick and S.J. Biggs (Eds.), *Using Video*. New York: John Wiley & Sons.

TALBOT, M. (1986). Speech recognition: is it working? *British Telecom Technology Journal*, **4**(2), 62-68.

TALBOT, M. (1987a). The future of speech technology. In J.A. Waterworth and M. Talbot (Eds.), *Speech and Language-based Interaction with Machines: Towards the Conversational Computer*. Chichester: Ellis Horwood.

TALBOT, M. (1987b). Adapting to the speaker in automatic speech recognition. *International Journal of Man-Machine Studies*, **27**, 449-457.

TAYLOR, M.M. (1986). Layered protocols in voice interaction with computers. Paper presented at the AGARD Aeromedical Panel Symposium on Information Management and Decision-Making in Advanced Airborne Weapons System, April 14-18, 1986, Toronto.

THIMBLEBY, H. (1982). Character level ambiguity: consequences for user interface design. *International Journal of Man-Machine Studies*, 16, 211-225.

THOMAS, J.C., and CARROLL, J.M. (1984). The Psychological Study of Design. In N. Cross (Ed.), *Developments in Design Methodology*. Chichester: John Wiley & Sons.

TOMS, P. (1988). Standards and the process of design. *Design Studies*, 9(2), April, 115-122.

TOVEY, M. (1989). Drawing and CAD in industrial design. *Design Studies*, 10(1), 24-39.

TSANG, P.S., HART, S.G., and VIDULICH, M.A. (1986). The effects of display-control I/O, compatibility, and integrality on dual-task performance and subjective workload. Paper presented at AGARD Conference on "Information Management and Decision-making in Advanced Airborne Weapon Systems", Toronto, Canada, 14-18 April 1986.

UGHANWA, D.O. (1988). In search of design excellence. *Design Studies*, 9(4), October, 219-222.

UTLEY, B. (1988). IBM: into the 1990s. *PC User*, 1-14, March edition, 38-40.

VAISSIERE, J. (1985). Speech recognition: a tutorial. In F. Fallside and W.A. Woods (Eds.), *Computer Speech Processing*. Englewood Cliffs, New Jersey: Prentice-Hall.

VAN HOOF, J.A.R.A.M. (1982). Categories and sequences of behavior: methods of description and analysis. In K.R. Scherer and P. Ekman (Eds.), *Handbook of Methods in Nonverbal Behavior Research*. Cambridge: Cambridge University Press.

VAN DER HEIDEN, G.H., and GRANDJEAN, E. (1984). Ergonomic studies in computer aided design. *Behaviour and Information Technology*, 3(4), 341-346.

VELTHUIJSEN, H., LIPPOLT, B.J., and VONK, J.C. (1987). A parallel blackboard architecture. *Proceedings of The Third International Conference on Expert Systems*.

VIDULICH, M.A. (1988). Speech responses and dual-task performance: better time-sharing or assymmetric transfer? *Human Factors*, 30(4), 517-529.

WARREN, C.P. (1987). Operation and representation of control in blackboard models. A working paper on Alvey Project "A User Modelling Tool for MMI Design" MMI142, Ref. MMI142-UCL-WP6.0 2.

WATERWORTH, J.A. (1982). Man-machine speech "dialogue acts", *Applied Ergonomics*, 13(3), 203-207.

WATERWORTH, J.A. (1984). Interaction with machines by voice - human factors issues. *British Telecom Technology Journal*, 2(4), 56-63.

WATERWORTH, J.A., and TALBOT, M. (1987). *Speech and Language-based Interaction with Machines: Towards the Conversational Computer*. Chichester: Ellis Horwood.

WELCH, J.R. (1980). Automatic speech recognition - putting it to work in industry. *Computer*, 14, 65.

WHITEFIELD, A.D. (1986a). Human factors aspects of pointing as an input technique in interactive computer systems. *Applied Ergonomics*, 17.2, 97-104.

WHITEFIELD, A.D. (1986b). Constructing and Applying a Model of the User for Computer System Development: the Case of Computer Aided Design. Unpublished PhD Thesis, University of London, England.

WHITEFIELD, A.D. (1986c). An analysis and comparison of knowledge use in designing with and without CAD. In *Knowledge Engineering and Computer Modelling in CAD, Proceedings of CAD 86*. London: Butterworths.

WHITEFIELD, A.D. (1986d). Human-computer interaction models and their use in computer system design. In *Proceedings of the Third European Conference on Cognitive Ergonomics - ECCE3*, September, Paris.

WHITEFIELD, A.D., and WARREN, C.P. (1989). A blackboard framework for modelling designers' behaviour. *Design Studies*, 10(3), July, 179-187.

WHITFIELD, D., BALL, R.G., and BIRD, J.M. (1983). Some comparisons of on-display and off-display touch input devices for interaction with computer generated displays. *Ergonomics*, **26**(11), 1033-1053.

WICKENS, C.D. (1983). Processing resources in attention, dual task performance, and workload assessment. In R. Parasuraman and R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

WICKENS, C.D. (1984). *Engineering Psychology and Human Performance*. Columbus, Ohio: Charles E. Merrill.

WICKENS, C.D., SANDRY, D.L., and VIDULICH, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, **25**(2), 227-248.

WICKENS, C.D., VIDULICH, M., and SANDRY-GARZA, D. (1984). Principles of S-C-R compatibility with spatial and verbal tasks: the role of display-control location and voice-interactive display-control interfacing. *Human Factors*, **26**(5), 533-543.

WILKINS, D.C., CLANCEY, W.J., and BUCHANAN, B.G. (1987). Knowledge base refinement by monitoring abstract control knowledge. *International Journal of Man-Machine Studies*, 27, 281-293.

WILLIAMSON, D.T., and CURRY, D.G. (1984). Speech recognition performance evaluation in simulated cockpit noise. *Proceedings of Speech Tech '84, Voice I/O Applications Show and Conference*, New York, 99-102.

WILLIGES, R.C. (1987). The use of models in human-computer interface design, The Ergonomics Society Lecture 1987. *Ergonomics*, **30**(3), 491-502.

WILLIGES, B.H., SCHURICK, J.M., SPINE, T.M., and HAKKINEN, M.T. (1986). Using speech in the human-computer interface. In R.W. Ehrich and R.C. Williges (Eds.), *Human-Computer Dialogue Design*. Amsterdam: Elsevier Science Publishers.

YOUNG, R.M. (1985). User models as design tools for software engineers. Unpublished paper given at Alvey workshop, September 1985, Loughborough, UK.

LIST OF APPENDICES

APPENDIX 1.	VITAS Programs for Scoring and Data Processing	301
APPENDIX 2.	A complete list of Behaviour Knowledge Sources and their descriptions	303
APPENDIX 3.	Questionnaire for Observational Study	307
APPENDIX 4.	A Summary of Questionnaire Findings	312
APPENDIX 5.	ANOVA Summary : Effect of Separate and Combined Systems on Eye Gaze to Screens	320
APPENDIX 6.	Summary of System Specifications for Speech Recognisers	321
APPENDIX 7.	Summary of System Specifications for Graphics Tablets	322
APPENDIX 8.	Summary of System Specifications for 2D Draughting CAD Software	323
APPENDIX 9.	Introduction to Experimental System and AutoCAD commands	324
APPENDIX 10.	CAD Task for Optimisation Study	328
APPENDIX 11.	Drawing Procedures for CAD Task - Optimisation Study	329
APPENDIX 12.	Recording Form for Optimisation of Speech Recogniser	330
APPENDIX 13.	Recording Form for Optimisation of Graphics Tablet	332
APPENDIX 14.	Questionnaire for Optimisation Study	334
APPENDIX 15.	Configuration of Recording Equipment	335
APPENDIX 16	Training Task for Experiment 1	337
APPENDIX 17.	Experimental Tasks for Experiment 1	338
APPENDIX 18a.	Training Overview and Instructions	340
APPENDIX 18b.	Subject Profile Form for Experiments 1 to 3	341
APPENDIX 19.	Recording Form for Experiments 1 and 2	342
APPENDIX 20.	Questionnaire for Experiment 1	343
APPENDIX 21.	ANOVA Summary for Experiment 1 Findings	345

APPENDIX 22.	A Summary of Questionnaire Findings - Experiment 1	350
APPENDIX 23.	Training Task for Experiment 2	352
APPENDIX 24.	Experimental Tasks for Experiment 2	353
APPENDIX 25.	Questionnaire for Experiment 2	355
APPENDIX 26.	ANOVA Summary for Experiment 2 Findings	358
APPENDIX 27.	Tablet Menu Overlay for Experiment 3	366
APPENDIX 28a.	Training Task for Experiment 3	367
APPENDIX 28b.	Experimental Tasks for Experiment 3	368
APPENDIX 29.	Speech List for Experiment 3	370
APPENDIX 30a.	Training Overview and Instructions for Experiment 3	371
APPENDIX 30b.	Recording Form for Experiment 3	372
APPENDIX 31.	Specific Instructions for Experiment 3	374
APPENDIX 32a.	Questionnaire for Systems F/G - Experiment 3	376
APPENDIX 32b.	Questionnaire for System F/H - Experiment 3	377
APPENDIX 33.	ANOVA Summary for Experiment 3 Findings	380
APPENDIX 34.	Proposed Initial Set of Guidelines for Integrating Speech and Manual Input in CAD Systems	388
APPENDIX 35a.	An Illustration of Guidelines Development Process	396
APPENDIX 35b.	Interview Schedule for Guidelines Validation Study	397
APPENDIX 36.	Suggestions by System Designers for Guidelines Redesign	398

APPENDIX 1. VITAS Programs for Scoring and Data Processing

Function	*Program Name*	*Data File Used*	*Data File Created*	*Information on Printout*
Scoring	EYES HAND TALK	- - -	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz (x=Subject no.) (y=Tape ID no.) (z=Analysis no.)	Raw data with time and behavioural codes
Reorganise hand movements into right, left and both	HAND TIMES	HAND Sx TyAz	HAND Sx HI	Each action is recorded with start and finish times
Identify anomalies in time and behavioural data sequences	CHECK FILES	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz	- - -	Prints action numbers where time codes are out of sequence or behavioural codes are repeated
Verify, identify or clarify any anomalies found in data files	CHECK TAPES	-	-	-
Edit time and behavioural codes, delete or insert actions, resolve anomalies from checking procedure	EDIT FILES	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz (modified)	- - -

[Appendix 1 continued]

Combines separate analyses	JOIN FILES	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz	EYES Sx TyAz HAND Sx TyAz TALK Sx TyAz	- - -
Condense subject scores	EYES MEANS	EYES Sx TyAz	EYES Sx MEANS	Absolute and % frequency and duration of eye movements
	HAND MEANS	HAND Sx HI	HAND Sx MEANS	Absolute and % frequency and duration of hand movements
	TALK MEANS	TALK Sx TyAz	TALK Sx MEANS	Absolute and % frequency of verbal content
	HAND TRANS	HAND Sx HI	HAND Sx TRANS	Absolute and % frequency and duration of hand transition
Produces group printout	GROUP EYES	EYES S- MEANS	-	Group eye means
	GROUP HAND	HAND S- MEANS	-	Group hand means
	GROUP TALK	TALK S- MEANS	-	Group talk means
	GROUP TRANS	HAND S- TRANS	-	Group hand transition means

APPENDIX 2. A complete list of Behaviour KSs and their descriptions

A brief description of each KS, with illustrative examples from behaviour protocols of expert and naive CAD users.

KS Name	*Description*
information :	command, coordinate, alphanumeric data and systems messages for executing systems function. eg. entering command.
command :	specific system instruction for specifying object entity and parameter. eg. entering command 'line' from screen menu.
graphical data : (coordinate)	single point on screen for positioning object entity. eg. pointing stylus tip to tablet, corresponding to its location on screen.
textual data : (text)	string of literal characters as annotation to drawing, object entity, etc. eg. depressing alpha keys on keyboard to enter text 'bathroom'.
numeric data : (numbers)	single number for specifying object parameter. eg. depressing numeric keys on keyboard to input numbers.
objects:	drawing objects as contained in plan such as chairs, charts, pipes, etc. eg. creating architectural object 'desk'.
instruction :	types of task procedures for doing the task such as speed, accuracy, etc. eg. specifying accuracy in the drawing.
menu overlays:	types of printed menus overlaid on tablet. eg. using menu overlay with engineering symbols
screen :	types of output screen for displaying graphics, alphanumeric and systems information such as graphics and text screens. eg. looking at graphics screen.
input device :	types of input devices for entering information such as graphics tablet, keyboard, speech recogniser, etc. eg. using graphics tablet to enter graphical data.
transducer :	types of probe devices used with input devices such as stylus, puck, etc. eg. pressing stylus tip to tablet surface.
task objects:	types of tools for aiding task such as CAD reference manual, drawing plan, map, calculator, etc. eg. holding plan to check object information.

[Appendix 2 continued]

elsewhere :	types of "non-task" objects such as colleagues, pen, pencil, drinking mug, etc. eg. talking to colleague
graphics screen :	displays graphics on graphics screen. eg. looking at drawing on graphics screen.
text screen :	displays text and systems information on text screen. eg. looking at system prompts on text screen.
graphics tablet :	manual input device with transducer and menu overlay . eg. looking at tablet overlay.
keyboard :	manual input device with alphanumeric, function and cursor keys. eg. looking at alphanumeric keys on keyboard.
speech recogniser :	voice input device fitted with headset microphone. eg. positioning the recogniser-microphone.
stylus :	pen-like transducer with button and sensitive tip. eg. holding stylus like a pen.
puck :	rectangular-shape transducer with multiple buttons and cross-shape cursor. eg. holding puck with finger on button.
graphics prompts:	system messages/feedback appearing on graphics screen. eg. looking at prompts on graphics screen.
text prompts:	system messages/feedback appearing on text screen. eg. looking at prompts on text screen.
tablet menu:	menu of commands and symbols overlaid on tablet surface. eg. pointing to 'valve' symbol in tablet menu.
screen menu:	menu of commands and symbols appearing on screen in the form of pull-down, pop-up or border menus. eg. looking at menu items in menu on graphics screen.
menu item :	data items in menu for specifying object. eg. selecting symbol 'screw' from menu.
puck key :	function or command buttons on puck. eg. depressing 'reset' key on puck.
function key :	button with associated command/task function(s). eg. pressing 'F1' key on keyboard.
alphanumeric key :	button with associated alphabet or number characters. eg. pressing character 'H' on keyboard.

[Appendix 2 continued]

calculator button :	buttons on calculator for inputting data. eg. depressing buttons on calculator.
drawing plan : (plan)	off-line drawing of object prototype showing arrangement and details of parts. eg. pointing to objects in a large-scale plan.
manual :	printed information concerning application program. eg. checking instruction manual on 'hatch' command.
offline vocabulary : list	off-line listing of words and phrases eg. holding list of spoken words.
online vocabulary : list	on-line listing of words and phrases eg. reading list of words on screen.
portable calculator :	hand-held device with multiple buttons for aiding mathematical calculations or programming. eg. holding calculator to calculate object dimensions.
computer calculator:	computer tool for aiding numerical calculations on screen. eg. looking at calculator on screen.
pencil/pen :	tool for writing. eg. holding pencil to write on plan.
colleague :	work associates or personnel. eg. talking to colleague in the office.
sense receptor :	types of sense organs for processing input and output. eg. using fingers to key-in data.
eyes :	eye movements to specific target. eg. eyes looking at graphics screen.
hands :	handedness of movements, one-hand or both-hands to target. eg. both hands transiting to keyboard.
right-hand :	right-hand movements to target. eg. using right hand to operate stylus.
left-hand :	left-hand movements to target. eg. using left hand to hold manual.
both-hands :	both-hand movements to target. eg. using both hands to hold up plan.
fingers :	fingers involvement in movements, some or all eg. some fingers keying-in data.

[Appendix 2 continued]

some-fingers :	few fingers involvement in movements. eg. using three fingers to press puck keys.
all-fingers :	all fingers involvement in movements. eg. using ten fingers to type-in data.
ears :	auditory processing of input. eg. ears listening for error feedback in the form of a ringing tone.
voice :	reproducing speech utterances. eg. voice verbalizing word 'listen'.
head :	orientation of head during movements. eg. head turning at an angle.
body :	parts of body to which movements are directed. eg. scratching legs.

APPENDIX 3. Questionnaire for Observational Study of CAD Experts at Work

Date given: Subject no.
Date back: Location:

QUESTIONNAIRE

Instructions:

Please answer all questions. Your answers should be based on the system in use and task performed on the day you were observed. Some questions would require you to:-

a. fill in the blanks
b. place a tick (/) in the spaces provided. Sometimes more than one answer is required.
c. delete whichever is not applicable, in particular those marked with an asterisk (*).

A. General Information
1. Age :
2. Job title :
3. Highest Qualifications :
4. Experience without CAD : * months/years
5. Experience with CAD :* months/years
6. Experience with current CAD system : * mths/years
7. No. of hours per day at CAD :
8. No. of days per week at CAD :
9. Handedness : Left
Right
Ambidextrous

B. Systems Information
i. Hardware

10. Type of Processor :
Micro (PC)
Mainframe
Mini
Others
11. Name of System :
Intergraph
Calma
Other
12. No. of terminal : One [Go to Qn. 13]
Two [Go to Qn. 12a-b]

a. If TWO, description of each terminal
separate graphics/text
combined graphics/text
others
b. Are both terminals used actively at a given time?
Yes
No
Sometimes
13. Type of Keyboard layout : QWERTY
other
14. Any function keys on keyboard?

[Appendix 3 continued]

Yes [Go to Qn. 14a]
No [Go to Qn. 15]

a. If YES, are the function keys in use?
Yes [Go to Qn. 14c] No [Go to Qn.14b]

b. Why not?

c. Most frequently-used function keys (please name, eg. F2 = reject, etc)

15. Do you use the keyboard more than the digitiser?
Yes
No
Sometimes

16. Type of device for tablet : stylus pen
puck/cursor

17. Any buttons on device? : No [Go to Qn. 18]
Yes [Go to Qn. 17a-d]

a. If YES, no. of buttons :
b. Function of buttons :
entering commands
drawing
making selection
others
c. Most commonly-used buttons :
d. No. of commands on puck :

18. Kind of menu overlay :
standard supplier's
tailor-made
combined
others

19. No. of menu overlays in use at any one time :

20. No. of commands on menu overlay :

21. How are menus organised? According to :
frequency of use
importance
others
don't know

22. Can menus be displayed on screen?
Yes [Go to Qn. 22a-b] No [Go to Qn. 23]
a. If YES, which screen?
text
graphics
both
b. Do you put menus on screen?
Yes [Go to Qn. 22c] No [Go to Qn. 23]

c. If YES, frequency of screen menu being used:
all the time/always
occasionally/sometimes
rarely/seldom

[Appendix 3 continued]

ii. Software/Program

23. Type of software : 2 D
3 D wireframe
3 D solid modelling

24. Name of Software :

iii. Context/Applications

25. Engineering : Process
Civil
Mechanical
Electrical
Mapping
Others

iv. Task

26. Nature of task : Creating new design
Editing existing drawing
Digitising * sketch/plan/map
Others

27. Nature of drawing : Detail
Layout

28. Design content (eg. piping, activity schedule,etc)

C. **Performance Information**

i. Commands

29. How are commands entered in the system? And how frequent are they entered via the various modes?

Frequency [always/sometimes/rarely]

Mode

Keyboard
Tablet menu
Screen menu
Button menu

30. Do you look at the command being entered/typed/selected via the following:-

Yes No always sometimes

Mode:

Keyboard
Tablet
Screen
Button

31. How do you know when you have entered/selected/typed-in a wrong command?

ii. Errors

32. Do you make errors while using the system?

Yes [Go to Qn. 32b] No [Go to Qn. 32a]

a. Why not?

b. If YES, what kind of errors? And why?

Due to/caused by +

Error Type: Yes No

1 2 3 4 5

typing
command
drawing
others

[Appendix 3 continued]

+

1 = having to look at the screen
2 = having to look away from the screen
3 = wrong selection/pointing
4 = mis-aiming/mislocating
5 = others (specify)

c. How do you know about errors?
error messages
beep sound [Go to Qn. 32e...]
visual feedback
others

d. Where do error messages appear?
text screen only
graphics screen only
both screens

e. Do you need to look at the screen
to recognise the error? Yes No
to correct the error? Yes No

f. How do you correct errors?
retype-in information
use error correction key on keyboard
use error button on puck
use specific command on menu overlay
others

iii. Prompts

33. Does the system provide prompts?
Yes [Go to Qn. 33a-e] No [Go to Qn. 34]

a. If YES, where are prompts given?
text screen only
graphics screen only
both screens

b. When are prompts given?

c. What sort of information do these prompts provide?

d. Do you read the prompts?
Yes, all the time
Yes, sometimes
No

e. Can you know what the prompts are without looking directly at them?
Yes How?
Sometimes
No

iv. Other information

34. Besides prompts and error messages, what other information do you receive from the system and
where are they displayed?
Screen type: Text/Graphics /Both
Information

D. **Ratings** [For these questions, put a slash mark on the line.
For example, good _____/____________ bad]

[Appendix 3]

35. What percentage of the total worktime at CAD terminal do you spend looking at screen(s)?

|---|

0 %   100 %

36. What percentage of total worktime at CAD terminal is spent on operating digitiser and keyboard?

|---|

0 %   100 %

37. How fluent are you in operating the digitiser/tablet?

|---|

Very fluent   Not fluent

38. How skilled are you in using the keyboard?

|---|

Unskilled   Very skilled

39. How do you rate your overall performance in using the system?

|---|

Satisfactory   Unsatisfactory
Why?

E. OPINIONS

40. Between the digitiser and keyboard, which do you prefer to use for the following functions:-

a. Entering commands? Why?
b. Drawing? Why?
c. Digitising? Why?
d. Inputting numeric data? Why?
e. Keying-in text? Why?

F. General Comments

Thank you for your time and effort.
Your willingness to participate in this study is greatly appreciated.

APPENDIX 4. A Summary of Questionnaire Findings: Observation of CAD Experts at Work

[Note: Numbers in brackets OR at the end of dotted line denote absolute frequency.]

A. Personal data

Variable	*Code*	*Number*
1. Age		
20 years and less	1	(3)
21 - 25	2	(2)
26 - 30	3	(2)
31 - 35	4	(2)
36 - 40	5	(1)
41 - 45	6	(2)
46 - 50	7	(2)
50 and above	8	(1)
2. Job Title		
cartographer	1	(3)
senior designer	2	(3)
design supervisor	3	(6)
technician trainee	4	(3)
3. Highest Qualifications		
Bachelor degree	1	(1)
HNC/diploma	2	(7)
ONC	3	(5)
O levels	4	(2)
4. Experience without CAD		
1 year and less	1	(2)
1 - 5	2	(3)
6 - 10	3	(2)
11 - 15	4	(1)
16 - 20	5	(2)
21 and more	6	(2)
No answer	0	(2)
Invalid answer	9	(1)
5. Experience with CAD		
1 year and less	1	(2)
1 - 2	2	(3)
3 - 4	3	(5)
5 - 6	4	(5)
6. Experience with current CAD system		
1 year and less	1	(5)
1 - 2	2	(1)
3 - 4	3	(6)
5 - 6	4	(3)
7. No. of hours per day at CAD		
6 hours and less	1	(1)
7-8	2	(7)
9-10	3	(6)
11 and more	4	(1)
8. No. of days per week at CAD		
3 days and less	1	(2)
4	2	(4)
5	3	(9)
9. Handedness		
Left	1	(4)
Right	2	(11)

[Appendix 4 continued]

Ambidextrous	3	(0)	

B. Systems Information

i. *Hardware*

10. Type of Processor		
Mini	1	(3)
Mainframe	2	(12)
11. Name of System		
Intergraph	1	(12)
Calma	2	(3)
12. No. of screens		
Two	1	(14)
Vague answer	9	(1)
a. Screen description		
separate graphics/text	1	(3)
combined graphics/text	2	(11)
no answer	0	(1)
b. Screens used actively at a given time		
yes	1	(12)
no	2	(0)
sometimes	3	(1)
no answer	0	(2)
13. Keyboard layout		
QWERTY	1	(15)
14. Function keys on keyboard		
yes	1	(15)
a. are they in use?		
yes	1	(13)
b. frequent keys		
Calma: (F1-10)		
c/c or X (F1)	1	(1)
reject or Y (F2)	2	(2)
repaint or Z (F3)	3	(1)
interrupt (F10)	4	(2)
Bechtel: (F1-14)		
match text (F1)	5	(3)
standard text (F5)	6	(2)
wt (F10)	7	(2)
xy (F12)	8	(3)
dl (F13)	9	(5)
ac (active cell)	10	(3)
lv	11	(1)
no	2	(2)
no answer	0	(1)
vague answer	9	(4)
c. why?		
BP:		
not loaded	1	(1)
not used in graphics	2	(1)

[Appendix 4 continued]

15. Keyboard more used than digitiser
    - yes 1 .... (2)
    - no 2 .... (8)
    - sometimes 3 .... (5)
16. Tablet input device
    - stylus pen 1 .... (3)
    - puck/cursor 2 .... (12)
17. Buttons on device
    - no 2 .... (3)
    - yes 1 .... (12)
    - a. no. of buttons
        - 12 1 ..... (11)
        - 4 2 .... (1)
    - b. functions
        - enter command 1 ..... (12)
        - select tentative point 2 ..... (9)
        - fix data point 3 ..... (9)
        - change command 4 ..... (4)
    - c. frequently-used button
        - command (C) 1 ..... (11)
        - data (D) 2 ..... (12)
        - reset (R) 3 ..... (9)
        - tentative (T) 4 ..... (8)
    - d. No. of commands on puck
        - 4 1 ..... (2)
        - 8 2 ...... (1)
        - 12 3 ...... (3)
        - vague answer 9 ...... (2)
        - no answer 0 ...... (7)
18. Kind of menu overlay
    - standard supplier only 1 .... (3)
    - combined - own and standard 2 .... (12)
19. No. of menus in use at any one time
    - 1 1 .... (6)
    - 2 2 .... (4)
    - 3 3 .... (4)
    - 4 and more 4 .... (1)
20. No. of commands on menu overlay
    - 200 and more 1 .... (5)
    - no answer 0 .... (4)
    - vague answer 9 .... (6)
21. Menu organisation
    - frequency of use 1 .... (8)
    - importance 2 .... (5)
    - topical 3 .... (3)
    - don't know 9 .... (2)
22. Menu displayed on screen
    - no 2 .... (5)
    - yes 1 .... (10)
        - a. which screen?
            - text (Screen 1) 1 ...... (0)
            - graphics (Screen 2) 2 ...... (0)
            - both 3 ...... (10)
        - b. use screen menu?
            - no 2 ...... (5)

[Appendix 4 continued]

	yes	1	(4)
	no answer	0	(1)
c. frequency of use			
	all the time	1	(0)
	occasionally	2	(4)
	no answer	0	(1)

ii. Software/Program

23. Type of software

2 D	1	....	(10)
3 D wireframe	2	....	(5)
3 D solid modelling	3	....	(5)

24. Name of software

Intergraph	1	....	(1)
Calma	2	....	(3)
Bechtel	3	....	(9)
no answer	0	....	(2)

iii. *Context/Applications*

25. Engineering

process	1	....	(1)
civil	2	....	(4)
mechanical	3	....	(2)
electrical	4	....	(1)
mapping	5	....	(2)
scheduling	6	....	(2)
structural	7	....	(4)

iv. *Task*

26. Nature of task

create new design	1	....	(6)
edit existing design	2	....	(8)
digitise-in map/plan	3	....	(1)

27. Nature of drawing

detail	1	....	(4)
layout	2	....	(11)

28. Design content

piping	1	....	(7)
equipment	2	....	(1)
steelwork structure	3	....	(1)
architectural layout	4	....	(2)
schedule	5	....	(2)
geographical feature	6	....	(1)
flow diagram	7	....	(1)
vague answer	9	....	(1)
no answer	0	....	(2)

C. Performance Information

i. *Commands*

29. Entry mode (and Frequency : 1=always; 2=sometimes; 3=rarely)

keyboard	1	....		(15)				
			1	(3)	2	(11)	3	(0)
tablet menu	2	....		(15)				
			1	(6)	2	(9)	3	(0)
screen menu	3	....		(9)				
			1	(0)	2	(6)	3	(3)
button menu	4	....		(7)				
			1	(4)	2	(2)	3	(1)

[Appendix 4 continued]

30. Looking at command entry (and Frequency : Yes, always=1; yes, sometimes=2; No=3)

keyboard	1	(15)						
		1	(11)	2	(3)	3	(0)	
tablet	2	(15)						
		1	(10)	2	(5)	3	(0)	
screen	3	(10)						
		1	(6)	2	(5)	3	(0)	
button	4	(6)						
		1	(4)	2	(2)	3	(0)	

31. Information source for command error

invalid prompt on screen	1	(8)
beep sound	2	(3)
expected result not shown	3	(3)
error messages	4	(3)
visual check	5	(2)

ii. *Errors*

32. Errors while using system

no	2	(0)
yes	1	(15)

a. Reasons

yes	1	(15)

b. Error kinds (and causes: 1=having to look at screen;2=having to look away from screen;3=wrong selection;4=mis-aiming/mislocating;5=others)

typing	1	(15)
		1 ....(1) 2 ....(4) 3 ....(1) 4 ....(6) 5 ....(3)
command	2	(15)
		1 ....(0) 2 .... (1) 3 .... (9) 4 ....(5) 5 ....(0)
drawing	3	(10)
		1 ....(0) 2 .... (0)3 ....(1) 4 ....(6) 5 ....(3)

c. Error information

error messages	1	(15)
visual feedback	2	(10)
beep sound	3	(6)
inexecutable command	4	(1)

d. Location of error information

text (screen 1) screen only	1	(3)
graphics (screen 2) only	2	(2)
both screens	3	(9)
no answer	0	(1)

e. Looking at screen:

- to recognise error

yes	1	(14)
no	2	(0)
no answer	0	(1)

- to correct error

yes	1	(9)
no	2	(4)

[Appendix 4 continued]

sometimes	3	(1)	
no answer	0	(1)	

f. Error correction method

retype-in information	1	(15
use function key on keyboard	2	(4)
use error button on puck	3	(3)
use specific command on overlay	4	(7)

iii. *Prompts*

33. System provide prompts

no	1	(0)
yes	2	(15)

a. location of prompts messages

text (screen 1) screen only	1	(3)
graphics (screen 2) only	2	(4)
both screens	3	(8)

b. when given

command initiated/selected	1	(10)
certain point in command sequence	2	(2)
special sub-programs	3	(2)
vague answer	9	(2)

c. type of information provided

select next input	1	(4)
reject previous input	2	(1)
system readiness	3	(2)
option selection	4	(3)
geometric/coordinate input	5	(4)
enter text	6	(1)
confirm command	7	(3)
vague answer	9	(2)

d. Need to read prompts

yes, all the time	1	(7)
yes, sometimes	2	(8)
no	3	(0)

e. Know prompts without looking

yes	1	(3)
sometimes	2	(9)
no	3	(2)
no answer	0	(1)

How?

nature of command input	1	(3)
memory/remembering	2	(3)
familiar with sequence	3	(3)
peripheral vision	4	(2)
no answer	0	(5)

iv. *Other Information*

34. What information and where displayed

systems messages		1	(10)
(eg. automatic reminder to save file;	text	1	(5)
query replies)	graphics	2	(1)
	both screens	3	(4)
no answer		0	(5)

[Appendix 4 continued]

D. Ratings

35. Percentage of total worktime at CAD terminal - looking at screens
36. Percentage of total worktime operating controls
37. Fluency in operating tablet/digitiser (%)
38. Skill in keyboard use (%)
39. Overall performance in system use (%)

Subject	Qn.35	Qn.36	Qn.37	Qn.38	Qn.39
1	90.0	95.0	92.0	66.0	88.0
2	72.0	53.0	83.5	58.5	96.5
3	84.5	13.0	85.0	73.0	91.0
4	66.5	98.0	86.0	30.0	84.0
5	69.0	20.0	78.0	58.0	78.0
6	64.0	17.0	56.5	70.5	84.0
7	97.0	99.0	97.0	49.0	99.9
8	81.5	47.5	94.0	87.0	94.0
9	69.0	45.0	86.0	67.0	75.0
10	53.0	29.0	93.0	70.0	99.9
11	83.0	38.5	79.0	51.5	89.5
12	68.5	70.0	89.5	49.0	87.5
13	74.5	19.0	60.5	63.0	21.0
14	94.0	68.0	88.0	42.0	83.0
15	60.0	46.0	41.0	31.5	51.5

39. Reasons for performance rating

still learning/training incomplete	1	3
familiarity with system	2	3
experience/constant use	3	4
great improvement in task	4	1
work completion on time	5	1
no answer	0	4

E. Opinions

40. Preference of input device and reasons

a. entering commands

tablet	1	11
keyboard	2	3

speed/quicker	1	6
simple/easier	2	2
convenient	3	2
less memory work	4	1
only method	5	0
no utility on device	6	0
accuracy	7	0
rewarding	8	0
vague answer	9	1
no answer	0	4

b. drawing

tablet	1	11
keyboard	2	3

speed/quicker	1	3
simple/easier	2	2
convenient	3	2
less memory work	4	1

[Appendix 4 continued]

			only method	5	.....	1
			no utility on device	6	.....	0
			accuracy	7	.....	1
			rewarding	8	.....	0
			vague answer	9	.....	2
			no answer	0	.....	4

c. digitising

tablet	1	......	12
keyboard	2	......	2

speed/quicker	1	.....	1
simple/easier	2	.....	0
convenient	3	.....	1
less memory work	4	.....	1
only method	5	.....	4
no utility on device	6	......	0
accuracy	7	.....	1
rewarding	8	.....	0
vague answer	9	.....	1
no answer	0	.....	6

d. inputting numeric data

tablet	1	......	0
keyboard	2	......	15

speed/quicker	1	.....	2
simple/easier	2	.....	2
convenient	3	.....	0
less memory work	4	.....	0
only method	5	.....	4
no utility on device	6	.....	2
accuracy	7	.....	2
rewarding	8	.....	0
vague answer	9	.....	0
no answer	0	.....	3

e. keying-in text

tablet	1	......	0
keyboard	2	......	15

speed/quicker	1	.....	2
simple/easier	2	.....	3
convenient	3	.....	0
less memory work	4	.....	0
only method	5	.....	6
no utlity on device	6	......	1
accuracy	7	.....	1
rewarding	8	.....	0
vague answer	9	.....	0
no answer	0	.....	4

APPENDIX 5. ANOVA Summary : Effect of Separate (n=3) and Combined (n=12) Systems on Eye Gaze to Screens

1. Effect of System on Frequency of Eye gaze to Graphics screen (Screen 1)

Source	*Sum of squares*	*DF*	*Mean square*	*F*	*p*
System	54.51	1	54.51	1.64	.22
Residual	433.41	13	33.34		
Total	487.92	14	34.85		

2. Effect of System on Frequency of Eye gaze to Text screen (Screen 2)

Source	*Sum of squares*	*DF*	*Mean square*	*F*	*p*
System	172.69	1	172.69	4.14	.06
Residual	542.53	13	41.73		
Total	715.22	14	51.09		

3. Effect of System on Duration of Eye gaze to Graphics screen (Screen 1)

Source	*Sum of squares*	*DF*	*Mean square*	*F*	*p*
System	15.21	1	15.21	.10	.76
Residual	1996.83	13	153.60		
Total	2012.04	14	143.72		

4. Effect of System on Duration of Eye gaze to Text screen (Screen 2)

Source	*Sum of squares*	*DF*	*Mean square*	*F*	*p*
System	32.00	1	32.00	.24	.64
Residual	1771.28	13	136.25		
Total	1803.28	14	128.8		

APPENDIX 6. Summary of System Specifications for Speech Recognisers

System name:	*Pronounce 1.20*	*Vocalink SRB-LC*	*VoiceScribe 1000*	*Voicekey*	*Voicelink VRS PC-512*	*IBM Board*
Supplier :	Microphonics	Interstate Products	Cherry	Cadfax Systems	Roar Technology	R & D Technology
Vocabulary size:	256 words/phrases	400 words/phrases	1000 words/phrases	512 words/phrases	512 words/phrases	96 words/phrases
Utterance length:	.5 to 2 sec.	.2 sec.	.2 to 2 sec.	.2 to 2 sec.	ND	.5 sec
Keystroke:	255 per word	ND	ND	64 per word	ND	ND
Training:	single pass	3 times	4 times	ND	single pass	2 times
Recognition:	90%	90%	99.3%	ND	ND	95%
Memory:	68K of PC-RAM	64K of PC-RAM	ND	self-contained	ND	self-contained
Board dimension:	full-length card	half-length card	full-length card	full-length card	full-length card	full-length card
Application:	CAD, database, word processing	word processing, database management	inventory control, medical records	CAD, database, word processing	CAD, database, word processing	CAD, word processing, process control
Price:	£795	£495	£895	£950	£880	£350

Note: All systems operate on IBM PC-XT/AT and 100% compatibles; fitted with headset microphone; and are speaker-dependent continuous word recognisers.

ND = no data (information)

Source: What's New in Computing, March, 1987, pp. 31-32.

APPENDIX 7. Summary of System Specifications for Graphics Tablet

System name:	*Summasketch*	*Drawingboard*	*Graphics Table*	*Numonics*	*Digicad*	*LCA4*
Supplier:	Summagraphics	Calcomp	Cherry	Comtek	Kontron	Terminal Data System
Dimensions (mm):	406x412x20	ND	ND	527x410x22	635x457x25	421x418
Digitising area (mm):	292x292	305x305	384x290	400x300	452x300	305x305
Accuracy (mm):	0.6	0.635	0.5	0.25-0.125	0.025	0.25
Resolution (mm):	0.1	0.1	0.1	0.025	0.025	0.05
Price:	£495	£570	£550	ND	£885	£720

Note: All systems can be used with a puck and/or stylus.

ND = no data

Source: CADCAM International, June, 1987, pp. 53-56.

APPENDIX 8. Summary of System Specifications for 2D Draughting CAD Software

System name:	*AutoCAD*	*Superdraft*	*T-Square*	*Versacad*	*Cadplan*	*Robocad-PC*
Supplier:	Autodesk	Superdraft System	Norrie Hill	Versacad	Vistec Business System	Robocom
Menu/Command driven:	both	both	both	ND	menu	both
Text & Graphics:	yes	yes	yes	yes	yes	yes
Colour:	16	NA	NA	ND	15	ND
Memory:	384K	512K	384K	256K	320K	512K
Peripherals:	tablet, mouse, keyboard	tablet, keyboard	tablet, joystick, keyboard	tablet, joystick, keyboard	tablet, mouse, keyboard	tablet, mouse, keyboard
Price:	£2,500	£3,000	£4,250	£2,995	£2,200	£1,250

Note: All softwares run on MS/PC-DOS operating systems; and can be applied to most CAD applications.

ND = no data

NA = not available

Sources: PC User, April, 1985, pp. 106-118 and CADCAM International, April, 1986, pp. 102-104.

APPENDIX 9. Introduction to Experimental System and AutoCAD Commands

1. Study aim

- to examine the use of input devices in computer aided design (CAD).

2. System

- the system comprises both hardware and software, as follows:

2.1 *Hardware*

a. Processor: IBM PC-XT

b. Input devices:

- QWERTY keyboard - for entering alphanumeric text
- SUMMASKETCH tablet and stylus - for entering graphical data and commands
- PRONOUNCE speech recogniser - for entering commands and numerical data

c. Output devices:

- IBM colour graphics screen - for displaying graphics and status information
- SAMSUNG monochrome text screen - for displaying text, command line and prompts
- HEWLETT-PACKARD plotter - for plotting a hard copy of the drawing
- IBM graphics printer - for printing a hard copy of text document.

2.2 *Software*

- The application package is AutoCAD version 2.17.

3. Application

- The domain of application is Computer aided design.

4. CAD Task

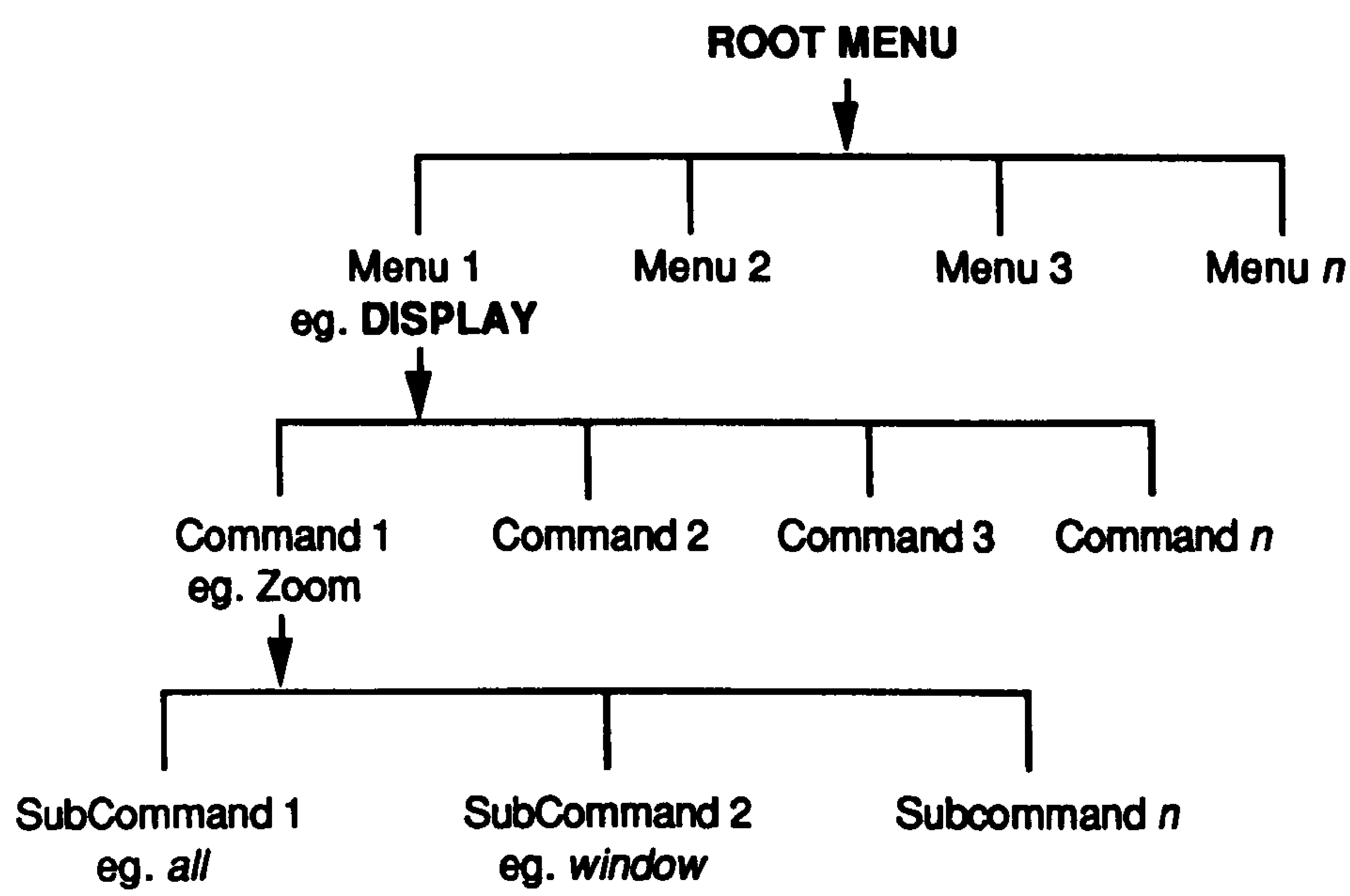
- The task involves modifying existing design plan and creating new objects, as shown in plan.

5. Content

- The design content is related to architectural work

[Appendix 9 continued]

AutoCAD Menu and Command Structure



[Appendix 9 continued]

AutoCAD Major Commands

ARC	[DRAW] Draws arcs of any size.
ARRAY	[EDIT] Makes multiple copies of selected objects in a rectangular or circular pattern.
AXIS	[MODE] Displays a "ruler line" on the graphics monitor.
BREAK	[EDIT] Erases part of an object or splits it into two objects.
CANCEL	[DOS] Cancels current command.
CHAMFER	[EDIT] Creates a chamfer at the intersection of two lines.
CHANGE	[EDIT] Alters properties of selected objects.
CIRCLE	[DRAW] Draws circles of any size.
COPY	[EDIT] Draws a copy of selected objects.
ENTER	[DOS] Registers input.
ELLIPSE	[DRAW] Draws an ellipse by INSERTing a circle, specifying x- and y-scale factors and axis orientation by dragging.
ERASE	[EDIT] Erases entities from the drawing.
FILL	[DRAW] Controls whether Traces are automatically filled on thescreen and the plot output.
FILLET	[EDIT] Constructs a smooth arc of specified radius between two lines.
GRID	[MODE] Displays a grid of dots, at desired spacing, on the screen.
HATCH	[HATCH] Performs cross-hatching and pattern-filling.
HELP	[UTILITY] Displays a list of valid commands and data entry options or obtains help for a specified command.
LAYER	[LAYER] Creates named drawing layers and assigns colour and linetype properties to those layers.
LIMITS	[UTILITY] Changes the drawing boundaries and controls checking of those boundaries.
LINE	[DRAW] Draws straight lines of any length.
LTSCALE	[LAYER] Specifies a scaling factor to be applied to all linetypes within the drawing.
MENU	[UTILITY] Loads a file of Drawing Editor commands into the menu areas.
MOVE	[EDIT] Moves designated entities to another location.

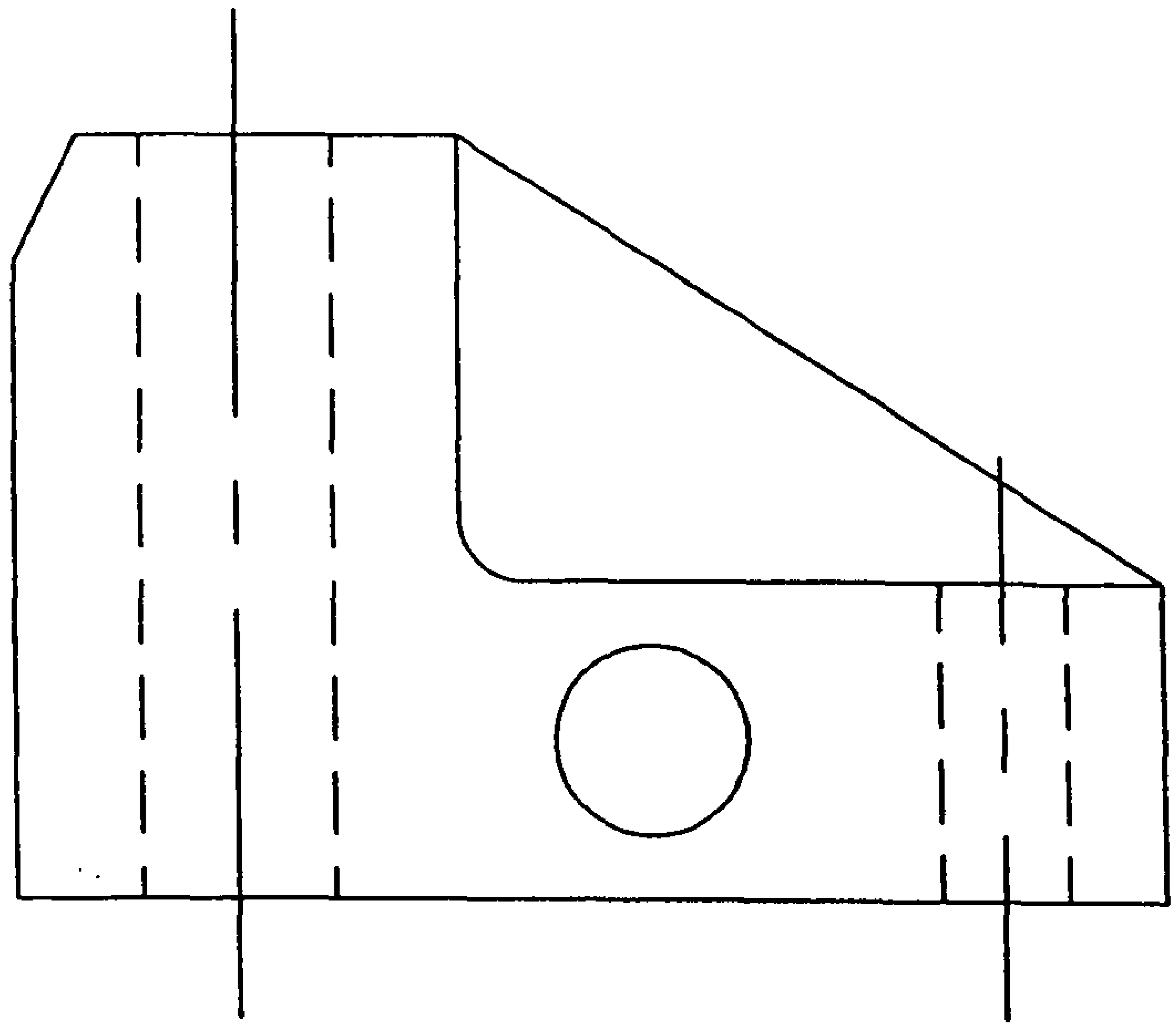
[Appendix 9 continued]

OOPS	[EDIT] Restores erased entities.
ORTHO	[MODE] Constrains LINE drawing so that only lines aligned with the current grid can be entered.
OSNAP	[MODE] Enables points to be precisely located on reference points of existing objects.
PAN	[DISPLAY] Moves the display window.
POINT	[DRAW] Draws single points.
QUIT	[UTILITY] Exits the Drawing Editor and returns to AutoCAD's Main Menu, discarding any changes to the drawing.
RECTANG	[DRAW] Draws a rectangle by INSERTing a square, specifying x- and y-scale factors and orientation by dragging.
REDRAW	[DISPLAY] Refreshes or cleans up the display.
REGEN	[DISPLAY] Regenerates the entire drawing.
SAVE	[UTILITY] Updates the current drawing file without exiting the Drawing Editor.
SNAP	[MODE] Specifies a "round-off" interval for tablet point entry so entities can be placed at precise locations easily.
STATUS	[INQUIRY] Displays statistics about the current drawing.
STYLE	[UTILITY/TEXT] Creates named text styles, with user-selected combinations of font, mirroring, obliquing, and horizontal scaling.
TABLET	[MODE] Configures the menu overlay for the tablet.
TEXT	[DRAW] Draws text characters of any size, with selected styles.
TRACE	[DRAW] Draws solid lines of specified width.
UNITS	[UTILITY] Selects coordinate and angle display formats and precision.
ZOOM	[DISPLAY] Enlarges or reduces the display of the drawing.

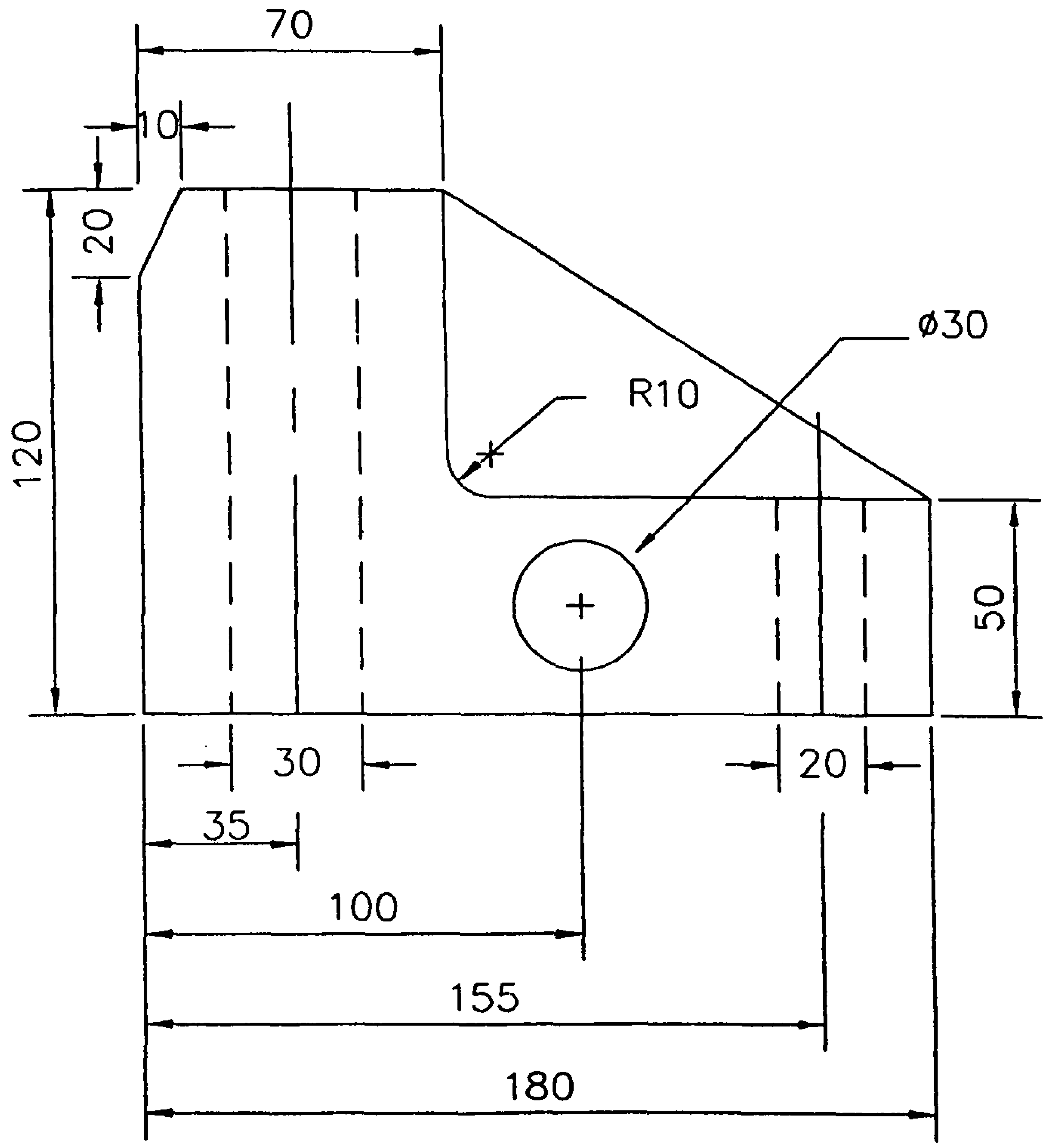
CAD Task for Optimisation Study

EX 1 STANCHION MOUNT

TARGET DRAWING



DRAWING INSTRUCTIONS



APPENDIX 11. CAD Task and Drawing Procedure for Optimisation Study

Task: Design a Stanchion Mount (as shown in plan).

Commands (as given below):

Mode	*Utility*	*Draw*	*Edit*	*Display*	*Layer*	*Dos*
Snap	Help	Line	Erase	Redraw	Layer	Enter
Grid	Limits	Circle	Change	Pan		Cancel
Osnap	Save	Arc	Copy	Zoom		
Ortho	Quit	Text	Move			
			Chamfer			
			Fillet			
			Break			
			Oops			

Drawing procedure:

Set drawing limits to 0,0 by 240, 180 (LIMITS).

Use ZOOM "all" to show the drawing limits on screen.

Set GRID to 5.

Set SNAP to 5 (ensure SNAP is on by checking the status line).

Draw the stanchion using LINE command.

Use FILLET command to add a fillet with radius of 10.

Use the CHAMFER command to add a chamfer with distances of 10 and 20.

Draw the circle with CIRCLE command.

Use the LAYER command to select "hidden" linetype for drawing lines across the object.

Use SAVE to store drawing on file.

Use QUIT to end the session. (Respond "yes" to prompt).

<u>Note</u>:
The remaining commands should be used where necessary (eg. to modify the elements during drawing).

APPENDIX 12. Recording Form for Optimisation of Speech Recogniser - PRONOUNCE

A. Optimise to guidelines

Item	Guidelines	Instructions	Comments
Headset microphone	adjust to fit head comfortably ease of use	examine flexibility,	
	position to the side of the mouth, NOT in front of it. 1.5" between microphone and corner of mouth	place thumb between microphone and mouth	
Voice level setting	permanently records information about normal speaking level	run SETLEVEL once. save as Your-name.	
	trains Pronounce to understand a sample vocabulary of 5 words		
	use natural, conversational level tone of voice. Do not over enunciate. Do not put hard "t's" and "d's" at the end of words. Do not click tongue or smack lips when saying a word.		
	words/phrases should be spoken within 500 msec to 2 sec in duration.		
Training	speak consistently. Pronounce will allow speech to vary by 20-30%. If not recognised, varied by > 30%.		
Recognition	in check mode, word with score >8 should be retrained. Problem areas: • words begin with P,B,V or E. • words end with t or d. • words too short (eg. No)		
	in check mode, check difficult word. Distance between first and second word should be at least 2 (eg. 5/7 : score 5 is near to recognised word; score 7 is near to difficult word).		
	train difficult word/phrases more than once using slightly different forms (eg. upper case/lower case).		
	check LISTEN (on) - a long beep; GOODBYE (off) - a high, then low tone; PRONOUNCE (to access vocabulary).		

[Appendix 12 continued]

B. Optimise to Human factors criteria

Criteria	Instructions	Observation
System accuracy		
• recognition	train and check test vocabulary	no. of errors - substitution - rejection - spurious
• single pass training	use words at DOS command	no. of single passes
Response speed (duration from stimulus to onset ie. word displayed to response offset ie. word spoken by subject)	display test vocabulary. Subject repeat word at DOS command. Each word done twice in random order.	record time (in secs.)
Ease of use		
• recall-speak		
• natural, conversational		
Flexibility		
• headset adjustment	check for comfort level	
• microphone positioning	check position	
• headset wire/cable	check head movement	
Other aspects		
• limb positioning	which finger(s) on stylus which hand(s) on keyboard measure sitting height	

C. Optimise to application task: CAD

The task involves designing a stanchion mount. Design activity will be recorded on video.

Starting time:

Ending time:

Quality of drawing:

Problem(s) experienced in using device:

APPENDIX 13. Recording Form for Optimisation of Graphics Tablet - SUMMASKETCH

A. Optimise to Guidelines

Item	Guidelines	Instructions	Comment
Stylus	emits low intensity magnetic field	examine stylus tip for emission	
	stylus must be within active area of tablet where grid is located for field to be sensed	check stylus-active area relationship	
	stylus and tablet need not be in contact	place half-inch book on tablet. Check effectiveness.	
	hold stylus like holding a pen, but perpendicular to tablet	examine ease of use	
	two switches on stylus.Press barrel switch to activate. Press tip to tablet to activate internal switch.	check switch-effect.	
Tablet	flat or tilted position.	examine default tilted positions: - high tilt - mid tilt - low tilt	

B. Optimise to Human factors criteria

Criteria	Instruction	Observation
System accuracy		
• pointing/selection	point to test commands in tablet menu	no. of errors: - misaiming - substitution - other
• single pass pointing	select same commands in different order	no. of single passes no. of repeated selections
Response speed (duration from stimulus onset ie. word displayed to response offset ie. stylus press)	select command when displayed. Same word repeated twice in random order.	record time (in secs.)
Ease of use		
• look-point-press	verify	
• natural pen-holding	verify	

[Appendix 13 continued]

Flexibility

• tablet height	check tilt position preferred
• stylus cabling	check hand movement

Other aspects

• limb positioning	which finger(s) used to hold stylus
	which hand(s) for typing
	measure sitting height

C. **Optimise to application task: CAD**

The task involves designing a stanchion mount. Design activity will be recorded on video.

Starting time:

Ending time:

Quality of drawing:

(s) experienced in using the device:

APPENDIX 14. Questionnaire for Optimisation Study

A. General information

1. Sex: Male/Female

2. Age: years

3. Ethnicity:

4. Spoken languages:

5. Handedness: Left/Right/Ambidextrous

6. Vision: Unaided /Aided -> spectacles .... contact lens ....

7. Computer experience: weeks/ months/ years

B. Opinion

8. Are you satisfied with using the devices to perform the design task?

   Yes .... No ....

9. Indicate your satisfaction by putting a slash mark (/) on the line below.

   Graphics tablet

   |---|

   dissatisfied very satisfied

   Speech recogniser

   |---|

   very satisfied dissatisfied

10. What are the major difficulties/problems with using the:

    Graphics tablet

    ....

    Speech recogniser

    ....

11. What are the strengths of the:

    Graphics tablet

    ....

    Speech recogniser

    ....

12. Given a choice, which input device would you most prefer to use for performing a CAD task?

    Graphics tablet .... Speech recogniser ....

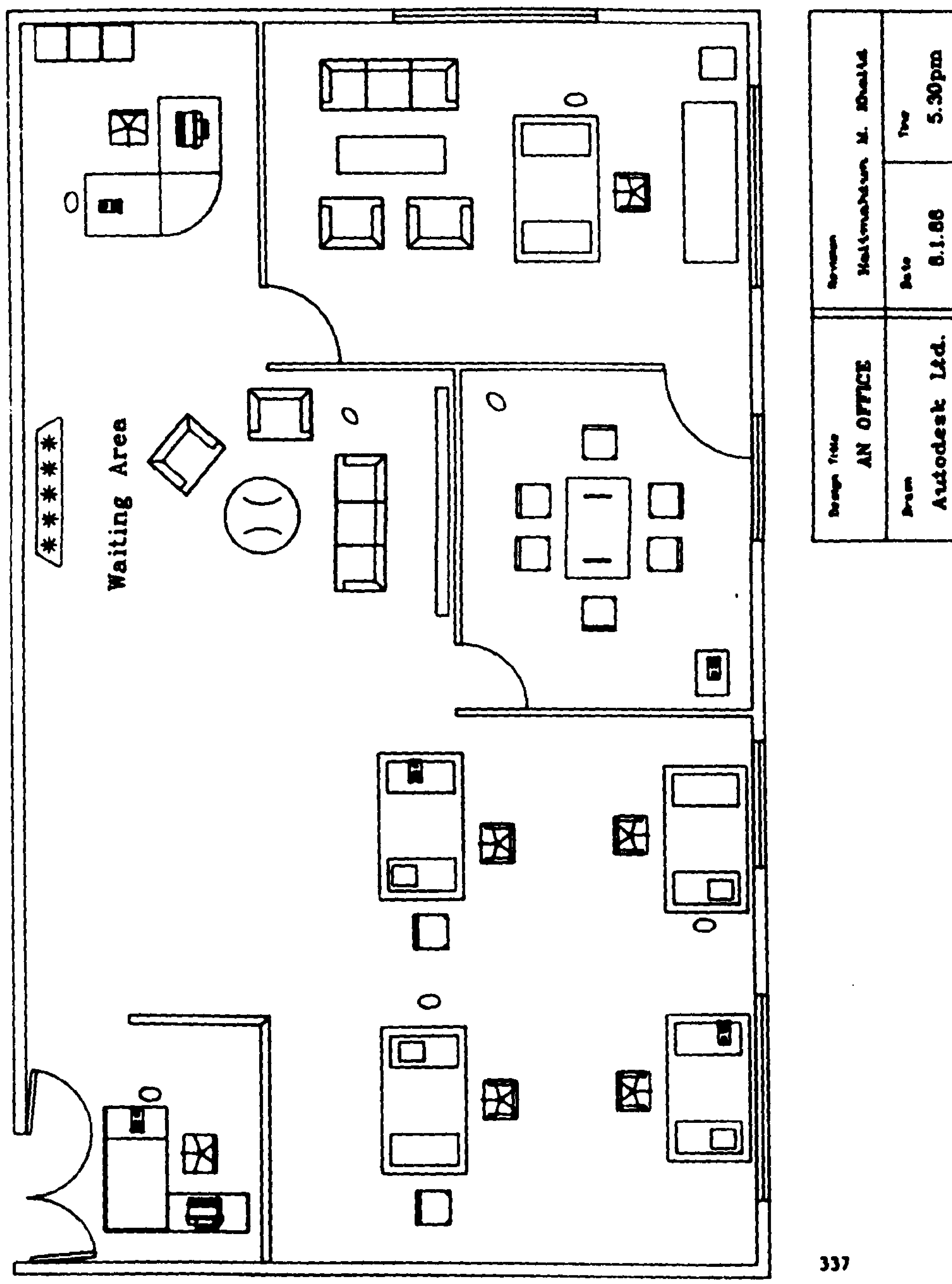
    Why?

Thank you for completing this questionnaire andfor participating in this study.

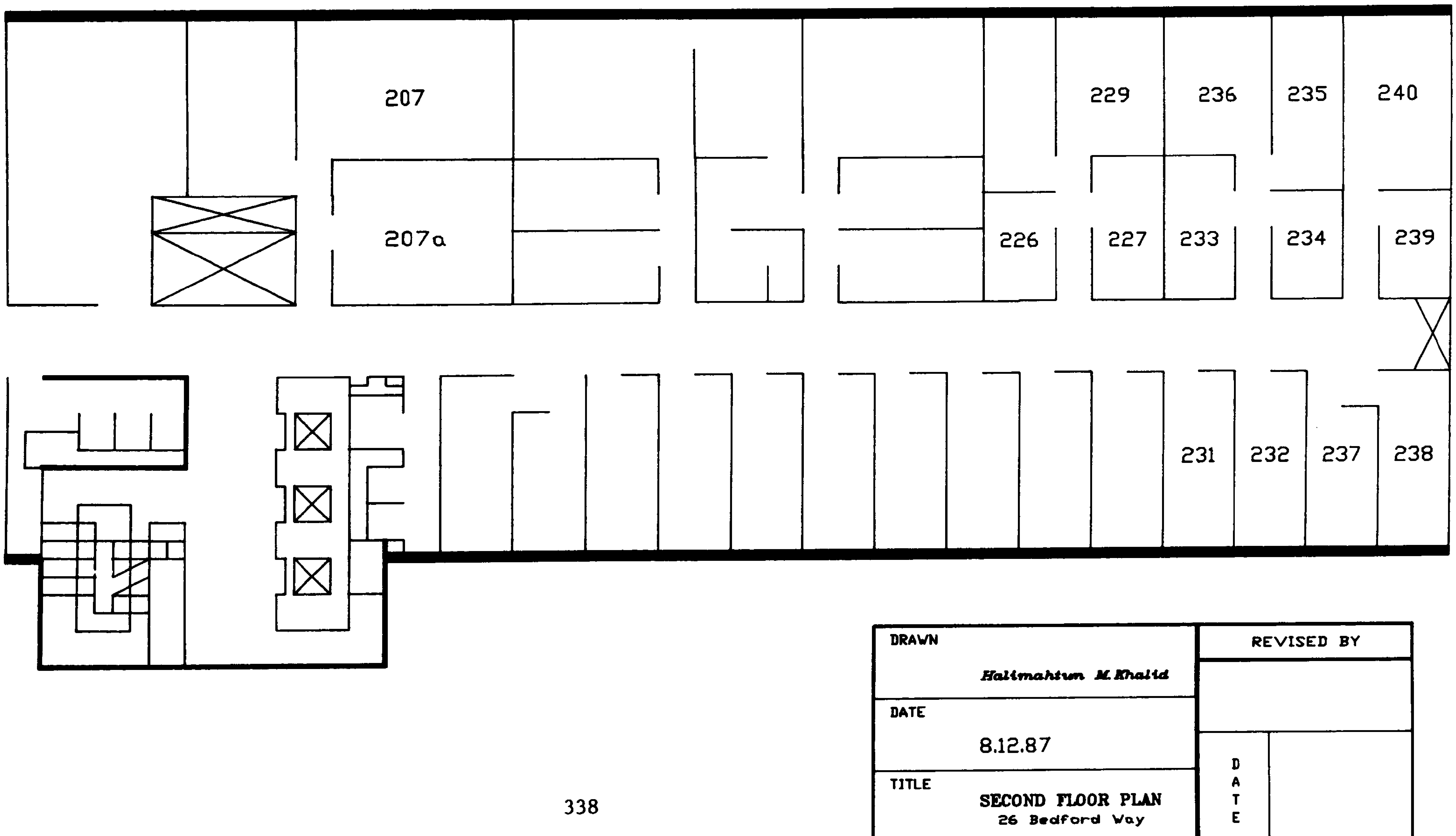
APPENDIX 15. Configuration of Recording Equipment

APPENDIX 15. Configuration of Recording Equipment

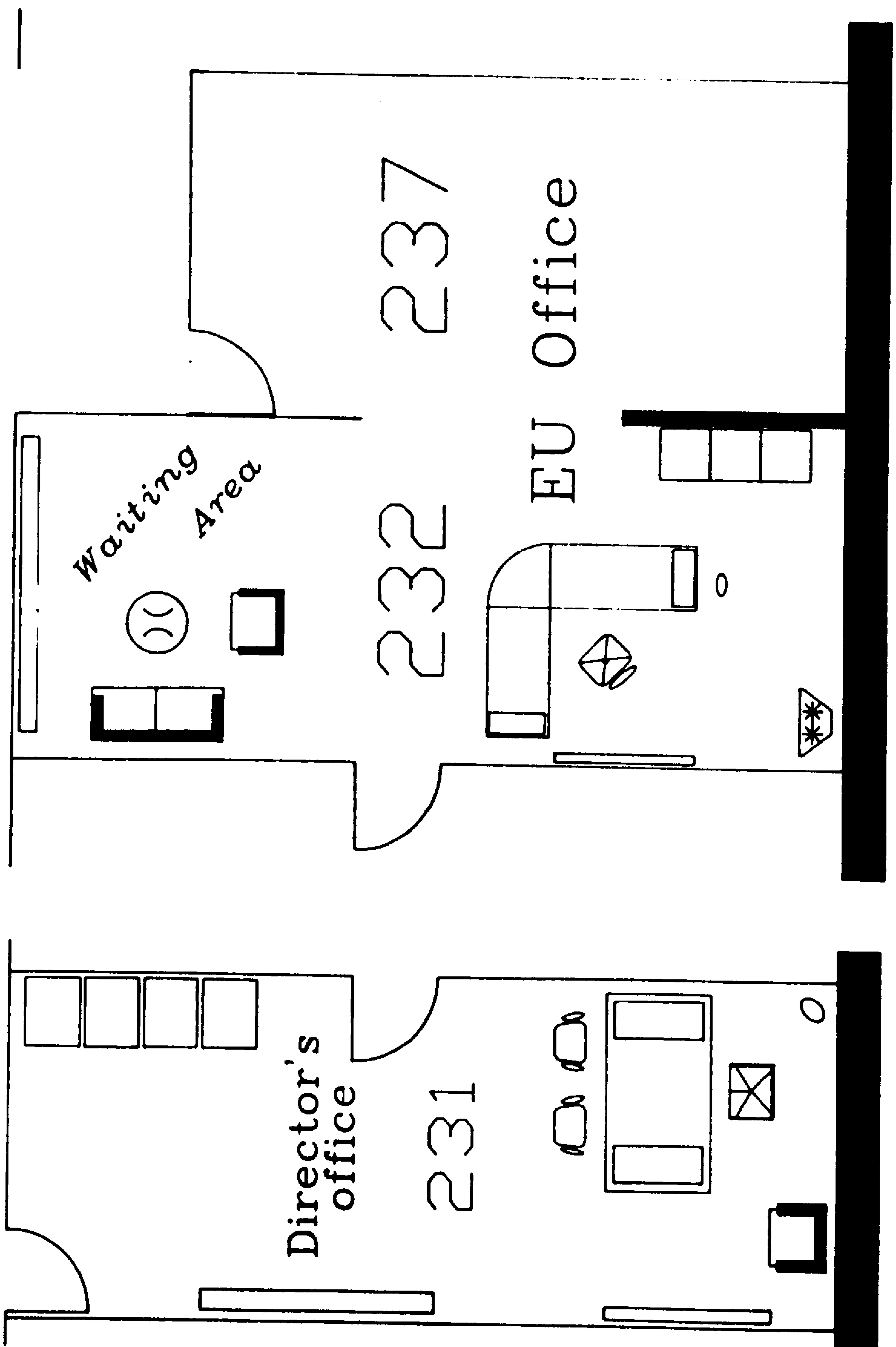
APPENDIX 16. Training Task for Experiment 1



APPENDIX 17. General office layout for Experimental tasks (Plans B and C)



APPENDIX 17. Experimental Tasks for Experiment 1 (Plan B and Plan C)



APPENDIX 18a. TRAINING : Overview and Instructions

1. INTRODUCTION

You are participating in an experiment which attempts to explore the use of speech and manual input devices for performing CAD tasks. This experiment is a part of my PhD research project. I am particularly interested in observing how you use these devices to perform simple design tasks. Your actions will be video-filmed to enable analysis of device usage. So, this is NOT a test of your performance nor skills.

You will first be trained on the application software, that is, AutoCAD 2.17. The aim of this brief training is to show you how to design and manipulate objects on a computer screen in order that you may achieve the basic skills required for performing three simple design tasks. Once trained you should be able to do the tasks on your own. Therefore, it is very important that you pay careful attention to the demonstration. Please ask questions whenever in doubt.

2. SESSIONS

There are TWO sessions which spans TWO days. You need to return for the second session within 1-2 days.

DAY 1 : Training on CAD and Input Devices

Phase 1
- briefing on systems used (Experimenter)
- complete subject profile form (Subject)

Phase 2
- demonstrate AutoCAD functions using keyboard (Experimenter)

Phase 3
- practise Task 1 for 45 minutes using the keyboard and completing the task (Subject)

DAY 2 : Perform two tasks further with both devices

Phase 4
- train on tablet (Tablet group) OR speech recogniser (Speech group)

Phase 5
- practise using tablet/speech recogniser for 15-20 minutes (Subject)

Phase 6
- perform Task 2 (Subject)
- rest (about 5 minutes)
- perform Task 3 (Subject)

Phase 7
- complete short questionnaire (Subject)

3. INSTRUCTIONS

- Try to do as much as you can and as EXACTLY as shown in the plan. If possible, try doing Tasks 2 and 3 without any help from me. You may use the HELP facility on-screen or the Reference Manual. But try to minimise referring to the manual as this will only slow down your performance.
- When using SPEECH, try to be consistent, that is, say the word the same way as you "trained" it.
- Try to remember the words and their precise location on the tablet menu so that you will not have to search each time.
- Lastly, I may request you to stop at some point even if you have not finished drawing.

I hope you will enjoy learning AutoCAD using both speech and tablet input. Any questions?

APPENDIX 18b. Subject Profile Form for Experiments 1 to 3

Completion Date: **Subject No.**

A. BIODATA

1. Name :
2. Sex : Male /Female
3. Age : years
4. Nationality :
5. Spoken Languages : English/ Others (please specify)
6. Handedness : Left/Right/Ambidextrous
7. Vision : Unaided /Aided

B. QUALIFICATIONS AND EXPERIENCE

8. Current status : Student/Employed / Others (specify)

 | |
 Subject/dept. Job title/dept.

 ---------- -----------

9. Highest qualifications :

 O levels / A levels / diploma /degree

 |
 Bachelor
 Master
 Doctorate

10. Computer experience : Yes No

 |
 days/weeks/months/years

11. CAD experience : Yes No

 |
 days/weeks/months/years
 Software used:________

12. Participation in previous study/experiment: Yes No

 |
 Input devices used:
 Keyboard/ Tablet / Speech Recogniser

13. Reasons for participating in this experiment :

 a. If *first-time* participation:

 b. If *second-time* participation:

Thank you very much for the information.

APPENDIX 19. Recording Form for Experiments 1 and 2

Recording Form

Subject Name: Subject No.

A. Training Session [Date : ....]

		Starting time	Ending time
1.	Briefing :	....	....
2.	Demo :	....	....
3.	Practice : [Plan1]	....	....
4.	Voice (Speech group):	....	....
5.	Vocab. 1/2 :	....	....

Comments:

B. Drawing Information [Plan 1]

1. Entities :
   Final :
   Original:
   Difference:
2. Errors :
3. Task completion time:

C. Experimental Sessions [Date : ....] Tape ID :

FIRST / SECOND CONDITION : A / B / C
TASK TWO / THREE (Plan) : A / B / C

		Starting time	Ending time
1.	Practice : YES / NO	....	....
2.	Perform :	....	....

Problems:

1. Speech
2. CAD
3. Tablet
4. Others

Drawing Information:

1. Entities :
   Final:
   Original A / B / C:
   Difference
2. Errors :
3. Set time : minutes

APPENDIX 20. Questionnaire for Experiment 1

Completion date: **Subject name:**

QUESTIONNAIRE

Instructions:

Please answer ALL questions. Some questions would require you to put a slash mark (/) on the line provided, others require you to explain.

A. **Performance**

1. How would you rate your overall performance?

   |---|

   poor excellent

2. Are you satisfied with your performance?

   |---|

   very satisfied very dissatisfied

3. What difficulties or problems did you experience in performing the task?
4. How did you overcome those difficulties/problems?
5. Did you find the first task compared to the second task

   more difficult

   equally difficult

   less difficult

B. **Device usability**

6. How would you rate the device you used for performing the task?

   |---|

   excellent poor

7. Is the device easy to use?

   |---|

   very difficult very easy

8. Did you find the device easy to learn?

   |---|

   extremely easy extremely difficult

9. What problems did you experience in using the device?
10. Did using the device help to improve your task performance?
11. In what way(s) did it help?
12. In what way(s) did it worsen your performance?
13. What would you consider the STRENGTHS of the device you used?
14. What are the LIMITATIONS of the device in the context of your task performance?

[Appendix 20 continued]

15. What would you recommend to improve to improve the usability aspects of the device you used?

C. **AutoCAD Commands**

16. Did you find the software easy to learn?

| -- |

difficult easy

17. What difficulties did you experience in implementing the command words provided?
18. Were the commands easy to remember?

Yes	No
|	|
all (100%)-------	Why not?
most (75%)------	
some (50%)------	
few (25%)-------	

19. Which were the most FREQUENT commands you used for performing the tasks? Please CIRCLE the commands in List A.
20. Which were the difficult commands to implement/use? Please CIRCLE them in List B.
21. Was the training on AutoCAD sufficient to enable you perform the tasks?

Yes No

D. **General comments**

22. Please comment on any aspects of the task, device, training, etc. that has not been mentioned above.

List A/B (Q.19 and Q.20)

Q.19. Please CIRCLE fifteen (15) of the most FREQUENT commands you used.
Q.20. Please CIRCLE ten (10) of the most DIFFICULT commands to use.

ARC	HELP	SAVE
AXIS	INSERT	SNAP
BREAK	LAYER	STATUS
CANCEL	LIMITS	STYLE
CHAMFER	LINE	TEXT
CHANGE	MOVE	TRACE
CIRCLE	OOPS	UNITS
COPY	ORTHO	ZOOM
ENTER	OSNAP	
ERASE	PAN	
FILL	QUIT	
FILLET	REDRAW	
GRID	RENAME	

APPENDIX 21. ANOVA Summary - Experiment 1 : Unitary Speech Input versusUnitary Manual (Tablet) Input

A. BEHAVIOUR MEASURES:

Effect of Unitary Systems on Frequency of Eye gaze to:

<u>All targets</u>									
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.7628	.7628	5.6730	.0263	Speech	12	2.2584	.3981
Within groups	22	2.9580	.1345			Tablet	12	1.9019	.3322
Total	23	3.7208				Total	24	2.080	.4022
<u>Graphics screen</u>									
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0054	.0054	4.5636	.0440	Speech	12	.1755	.0346
Within groups	22	.0262	.0012			Tablet	12	.2056	.0345
Total	23	.0316				Total	24	.1906	.0371
<u>Text screen</u>									
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0141	.0141	8.4783	.0081	Speech	12	.1714	.0343
Within groups	22	.0365	.0017			Tablet	12	.1230	.0463
Total	23	.0506				Total	24	.1472	.0469
<u>Graphics tablet</u>									
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.1133	.1133	261.1452	.0000	Speech	12	.0023	.0027
Within groups	22	.0095	.0004			Tablet	12	.1397	.0293
Total	23	.1229				Total	24	.0710	.0731
<u>Keyboard</u>									
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0042	.0042	33.2225	.0000	Speech	12	.0295	.0146
Within groups	22	.0028	.0001			Tablet	12	.0032	.0062
Total	23	.0069				Total	24	.0164	.0173

[Appendix 21 continued]

Drawing plan

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	1	.0002	.0002	.3111	.5826	Speech	12	.0567	.0274
Within groups	22	.0108	.0005			Tablet	12	.0618	.0152
Total	23	.0109				Total	24	.0592	.0218

Effect of Unitary Systems on Duration of Eye gaze to:

Graphics screen

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	1	915.1350	915.1350	17.5114	.0004	Speech	12	40.1750	6.7661
Within groups	22	1149.7050	52.2593			Tablet	12	52.5250	7.6641
Total	23	2064.8400				Total	24	46.3500	9.4750

Text screen

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	1	5142.1538	5142.1538	139.6331	.0000	Speech	12	44.1417	6.6098
Within groups	22	810.1758	36.8262			Tablet	12	14.8667	5.4738
Total	23	5952.3296				Total	24	29.5042	16.0872

Graphics tablet

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	1	2671.2600	2671.2600	249.5604	.0000	Speech	12	.1750	.3079
Within groups	22	235.4850	10.7039			Tablet	12	21.2750	4.6166
Total	23	2906.7450				Total	24	10.7250	11.2419

Keyboard

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev
Between groups	1	16.5004	16.5004	3.3339	.0815	Speech	12	2.7417	1.4362
Within groups	22	108.8858	4.9494			Tablet	12	1.0833	2.7993
Total	23	125.3863				Total	24	1.9125	2.3349

[Appendix 21 continued]

<u>Drawing plan</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	5.3204	5.3204	.6255	.4375	Speech	12	8.3583	3.4582
Within groups	22	187.1292	8.5059			Tablet	12	9.3000	2.2478
Total	23	192.4496				Total	24	8.8292	2.8926

Effect of Unitary Systems on Frequency of Hand:

<u>Idling</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0392	.0392	69.2156	.0000	Speech	12	.0775	.0253
Within groups	22	.0125	.0006			Tablet	12	.1584	.0222
Total	23	.0517				Total	24	.1180	.0474

<u>Drawing</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0025	.0025	6.1394	.0214	Speech	12	.0800	.0179
Within groups	22	.0091	.0004			Tablet	12	.1005	.0225
Total	23	.0116				Total	24	.0903	.0225

<u>Entering data</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0019	.0019	12.2839	.0020	Speech	12	.0187	.0172
Within groups	22	.0033	.0002			Tablet	12	.0011	.0025
Total	23	.0052				Total	24	.0099	.0150

<u>Locating menu</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0162	.0162	120.6479	.0000	Speech	12	.0000	.0000
Within groups	22	.0030	.0001			Tablet	12	.0520	.0164
Total	23	.0192				Total	24	.0260	.0289

[Appendix 21 continued]

Effect of Unitary Systems on Duration of Hand:

Idling

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	4100.8433	4100.8433	62.8523	.0000	Speech	12	65.8925	7.2090
Within groups	22	1435.4065	65.2458			Tablet	12	39.7492	8.8613
Total	23	5536.2498				Total	24	52.8208	15.5147

Drawing

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	933.8785	933.8785	11.1309	.0030	Speech	12	29.2950	8.0778
Within groups	22	1845.7878	83.8994			Tablet	12	41.7708	10.1266
Total	23	2779.6663				Total	24	35.5329	10.9934

Entering data

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	10.7870	10.7870	8.1235	.0093	Speech	12	1.5692	1.5345
Within groups	22	29.2135	1.3279			Tablet	12	.2283	.5486
Total	23	40.0005				Total	24	.8988	1.3188

Locating menu

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	1795.7400	1795.7400	183.0408	.0000	Speech	12	.0000	.0000
Within groups	22	215.8332	9.8106			Tablet	12	17.3000	4.4296
Total	23	2011.5732				Total	24	8.6500	9.3520

[Appendix 21 continued]

B. Performance measures

Effect of Unitary Systems on:

Product quality

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	32.6667	32.6667	1.0105	.3257	Speech	12	13.7500	5.5288
Within groups	22	711.1667	32.3258			Tablet	12	16.0833	5.8381
Total	23	743.8333				Total	24	14.9167	5.6869

Production cost (time)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.1760	.1760	1.5146	.2314	Speech	12	8.1095	.3553
Within groups	22	2.5565	.1162			Tablet	12	8.2807	.3259
Total	23	2.7326				Total	24	8.1951	.3447

Production cost (efficiency)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	.0567	.0567	11.6541	.0025	Speech	12	.4466	.0734
Within groups	22	.1070	.0049			Tablet	12	.3494	.0659
Total	23					Total	24	.3980	.0844

User acceptability (performance rating)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	319.4156	319.4156	.8437	.3683	Speech	12	57.0741	16.9271
Within groups	22	8328.6749	378.5761			Tablet	12	49.7778	21.6939
Total	23					Total	24	53.4259	19.3908

User acceptability (satisfaction rating)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	1	359.5144	359.5144	.5725	.4573	Speech	12	67.2593	26.3325
Within groups	22	13814.4362	627.9289			Tablet	12	59.5185	23.7162
Total	23	14173.9506				Total	24	63.3889	24.8246

APPENDIX 22. Summary of Questionnaire Findings - Experiment 1

Problems of task performance (Question 3)

Speech subjects:

Device
- headset uncomfortable
- voice inconsistency
- speech recognition
- stylus upright positioning

AutoCAD/task
- difficult commands
- remembering command functions
- decision over best way to draw
- frequent eye and hand movements from one device to another
- parameter specification
- confirmation of an entry using RETURN

Tablet subjects:

Device
- stylus sensitivity

AutoCAD/task
- limited practice
- transfer of training from keyboard to tablet
- separation of system information on two screens
- remembering commands
- understanding command functions
- knowing how to do it

Strategies for overcoming problems (Question 4)

Speech subjects:
- relax so that voice is consistent
- retrain command
- avoid certain commands that are difficult (eg. substitute 'Arc' for 'Fillet')
- be patient
- redraw entity
- try to memorise commands
- practice
- request help from experimenter
- trial and error
- repeat word calmly
- resort to keyboard if repeat attempts failed
- used same command whenever possible

Tablet subjects:
- think before doing
- use easier commands in place of difficult ones
- practice
- refer to manual
- trial and error
- memorise commands
- careful with stylus

[Appendix 22 continued]

Assessment of Input Systems

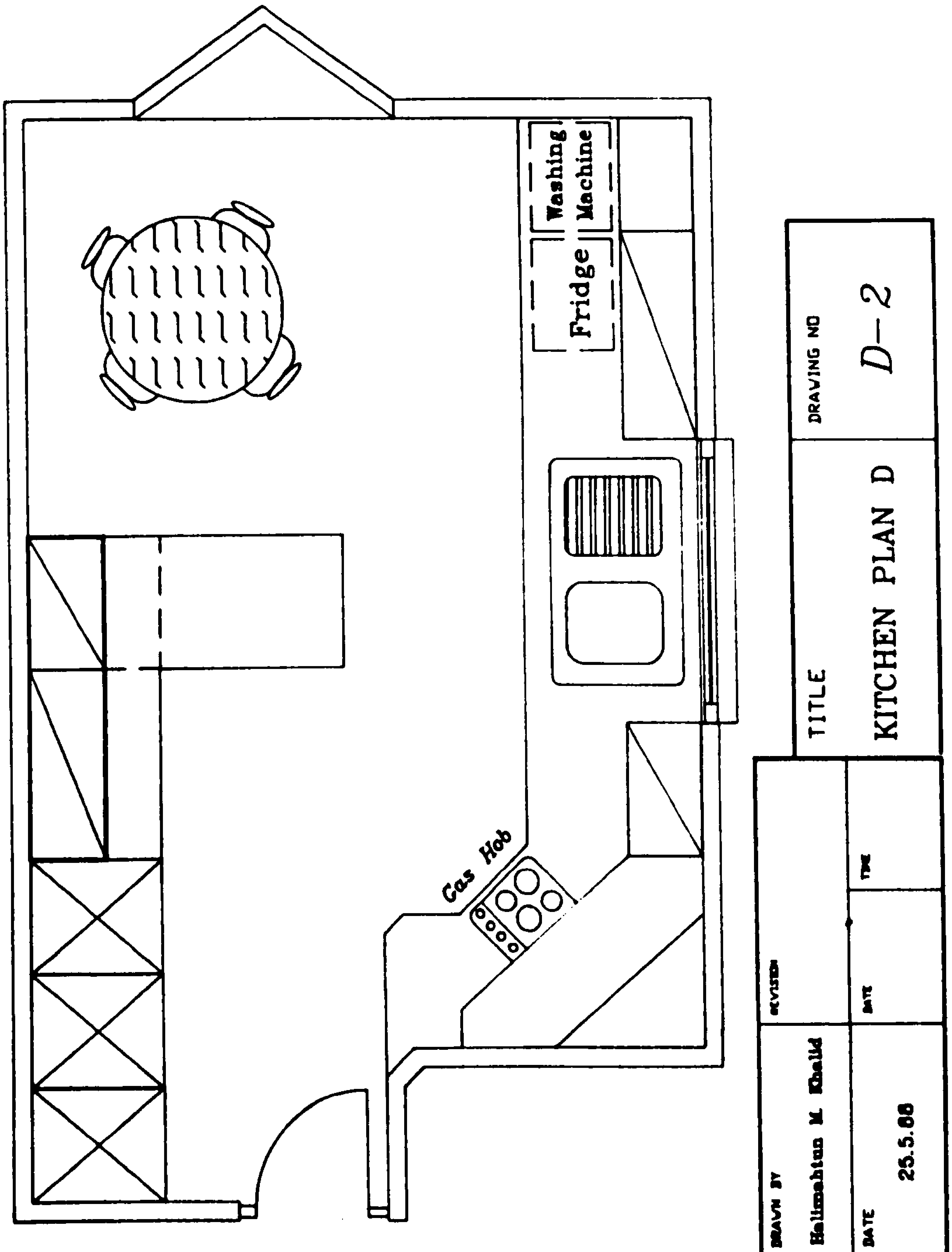
Speech subjects:
Strengths of recogniser
- attention less divided
- quicker than typing
- fun when recognised
- freedom of hands from keyboard
- less eye movement to tablet
- easy and pleasant to use
- more concentration on-screen and task
- natural
- speeds up task
- flexible

Limitations of recogniser
- background noise (eg. keyboard typing)
- necessity for retraining when voice changes (eg. cold)
- primed to desk because of headset cable
- speech recognition problems, especially similar-sounding words
- frustrating
- slow down performance due to problems
- time-consuming due to word repeat
- maintaining voice consistency
- inhibits chatting freely ("anti-social"), constrained from engaging in conversation
- limited vocabulary becuase of confusability
- rushed by the device to avoid confusability

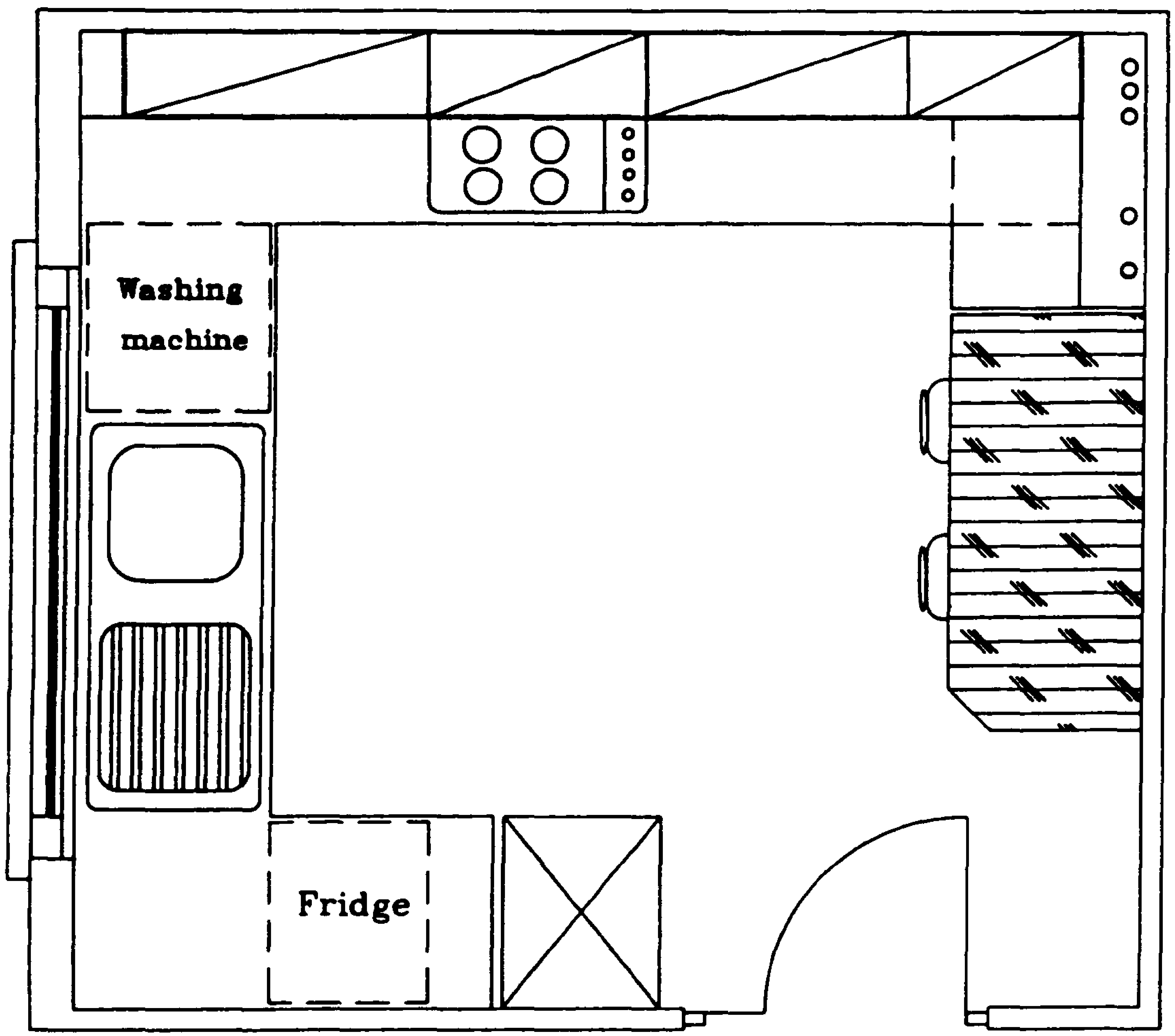
Tablet subjects:
Strengths of the tablet
- quick to use once concepts learnt
- clear display of commands on in menu
- quicker than typing
- well-organised menu overlay
- ease of use
- coordination of eye and hand
- easy to rectify errors
- no memory of commands

Limitations of tablet
- limited space for hand movements
- visual search of menu items, hence requires looking
- menu not large enough to include command details
- divided attention, look up and down
- boring, repetitive sequences

APPENDIX 23. Training Task for Experiment 2

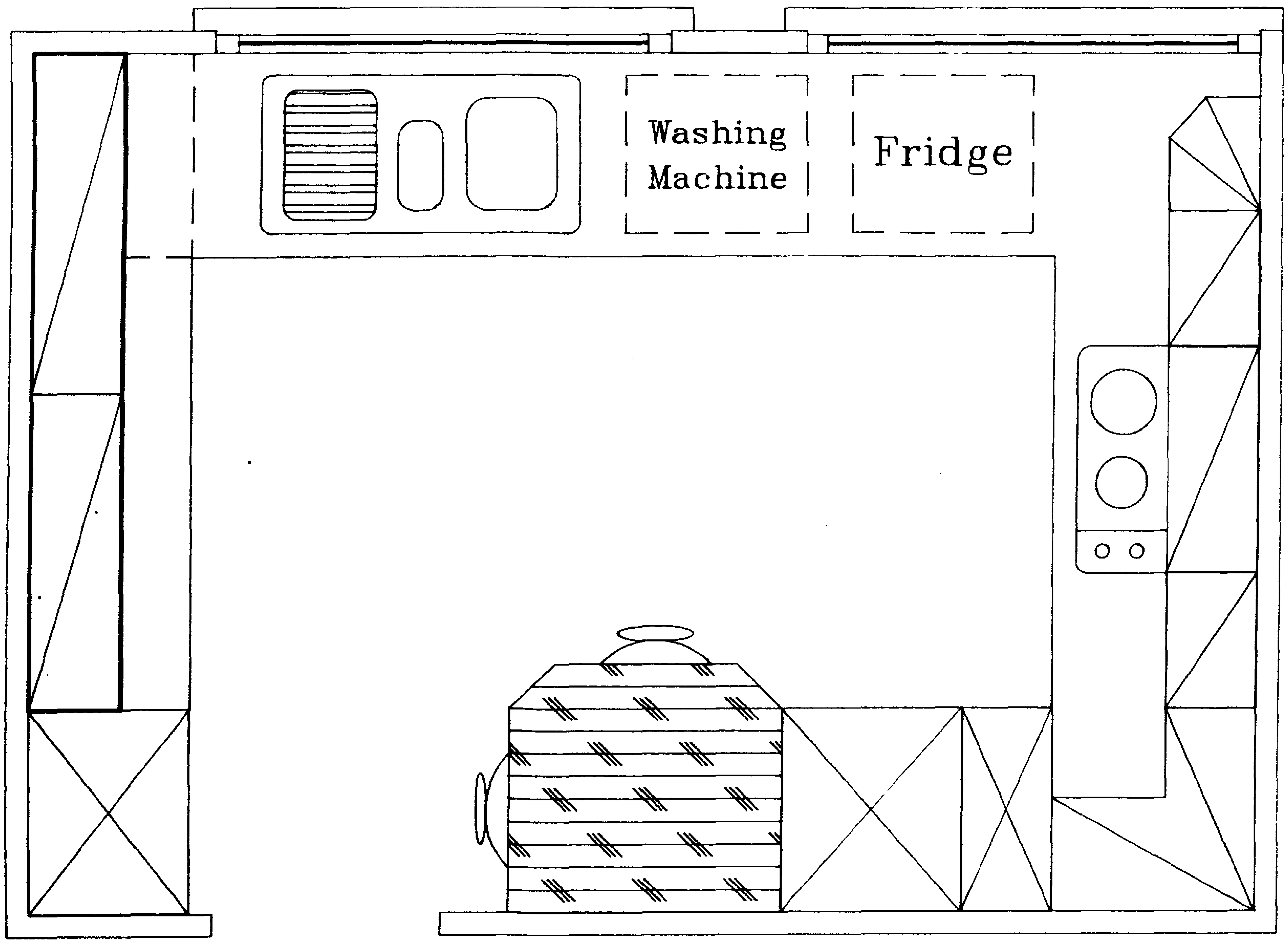


APPENDIX 24. Experimental Task for Experiment 2 (First plan)



EU DESIGNS	ERGONOMICS UNIT 26 Bedford Way London WC1H 0AP (01) 387 7050	
DRAWN BY Halimahtun M. Khalid		
REVISED	DATE	TIME
TITLE KITCHEN PLAN A		
DRAWING NO. A-2		

APPENDIX 24. Experimental Task for Experiment 2 (Second plan)



ERGONOMICS UNIT
26 Bedford Way
London WC1H 0AP
(01) 387 7050
DESIGNS

DRAWN BY
Halimahtun M. Khalid

REVISED

DATE

TIME

TITLE
KITCHEN PLAN B

DRAWING NO.
B–2

APPENDIX 25. Questionnaire for Experiment 2

Completion Date: Subject Name:

QUESTIONNAIRE

Instructions:

Please answer ALL questions. Some questions require you to put a slash mark (/) on the line provided, others require you to place a tick (/) against the relevant responses or elaborate.

1. How would you rate your overall performance in the second session?

   |---|

   excellent poor

2. Are you pleased with your performance?

   |---|

   not at all pleased very pleased

3. What difficulties or problems did you experience with the following. Put a check against those that apply to you. You may indicate as many as you experienced. Please consider each answer carefully...

3.1. Speech recogniser/Speech input

The following only refers to the Condition in which you used ALL SPEECH to enter commands/data.

a. Voice recognition: problem of being recognised
b. Vocabulary size: unmanageable
c. Remembering: difficult to remember commands
d. Meaning: problem of knowing/understanding meaning of command
e. Pronunciation: hard to pronounce some words
f. Consistency: difficult to remain consistent always
g. Confusion: misrecognition of commands creates confusion
h. Repetition: problem of having to repeat command
i. Weariness: speaking commands tires easily
j. Retraining: a lot of commands needed retraining
k. Activation: difficult to activate the system via mere "Listen" & "Goodbye"
l. Discomfort: headset microphone not comfortable
m. Constraint: headset cable constrained head movements

[Appendix 25 continued]

n. Usability: difficult to use the speech system
o. Learnability: device not easy to learn
p. Performance: device worsened task performance

3.1.1. Given the problems above, how did you overcome in order to perform the task? Please explain.

3.1.2. What are the advantages of using all speech input? Please elaborate.

3.2. *Speech and Tablet input*

The following only refers to the Condition in which you used both SPEECH and TABLET to enter commands/data.

a. Vocabulary size for Speech: unmanageable
b. Menu size for Tablet: unmanageable
c. Remembering: difficult to remember what had to be spoken and which had to be stylused
d. Performance: device-combination worsened task performance
e. Confusion: problem of identifying command-to-device mapping led to confusion
f. Coordination: difficult to coordinate speaking and pointing activities
g. Constraint: allocation of some words to speech and some to tablet constrained fluency in carrying out the task
h. Recall: difficult to recall some spoken words
i. Recall: problem of recalling location of some words in tablet menu
j. Delay: long lags between each drawing operation due to recall problem

3.2.1. Given the above problems, how did you overcome the difficulties?

3.2.2. What are the advantages of using dual-device combination for entering commands/data?

3.3. Preference and Solutions

3.3.1. Of the two conditions, ie. single mode (all speech) and dual mode (speech and tablet), which do you most prefer?

Single mode ....　　Dual mode ....

Why? Please explain.

3.3.2. Of the two single-mode conditions. ie. all speech and all tablet (eg. training session), which do you prefer?

All Speech ....　　All tablet ....

Why? Do explain.

3.3.3. In what ways do you think the dual-mode device usage can be further improved? That is, how can speech be combined with tablet input to better performance. Any ideas?

[Appendix 25 continued]

4. Task

4.1. Of the two tasks performed in the second session, was one task more difficult than the other OR both were equally difficult?

One more difficult .... [GO TO Qn. 4.2]

Both equally difficult .... [GO TO Qn. 4.3]

4.2. Which was more difficult?

First condition/Task 2 ....

Second condition/Task 3 ....

4.3. Did you find the tasks interesting? Yes .... No .... Don't know ....

4.4. Was the training on AutoCAD sufficient to allow you perform the tasks?

Yes .... No .... Don't know ....

5. Command/Data Usage

5.1. Frequent words in both conditions

Please tick the most FREQUENT words used to perform the design tasks. [Use Speech A list enclosed.]

5.2. What percentage of the total word list did you actually use for carrying out the tasks?

25%	(approx. 25-30 words)	....
50%	(approx. 50-55 words)	....
75%	(approx. 75-80 words)	....
100%	(approx. 100-105 words)	....

THANK YOU VERY MUCH FOR COMPLETING THIS QUESTIONNAIRE AND KINDLY PARTICIPATING IN THE EXPERIMENT.

APPENDIX 26. ANOVA Summary - Experiment 2: Integrated Speech-Manual versus Unitary Speech Input (n=24)

A. BEHAVIOUR MEASURES:

Effect of Integrated and Unitary Systems on Frequency of Eye gaze to:

<u>All targets</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	3.1421	1.5710	9.3435	.0004	System D*	12	2.1672	.2229
Within groups	45	7.5664	.1681			System E*	12	1.9692	.4135
Total	47	10.7085				System C	24	2.5604	.4727
						Total	48	2.3143	.4773

<u>Graphics screen</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0034	.0017	1.4388	.2479	System D	12	.1769	.0333
Within groups	45	.0531	.0012			System E	12	.1613	.0396
Total	47	.0565				System C	24	.1564	.0321
						Total	48	.1628	.0347

<u>Text screen</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0036	.0018	1.1606	.3225	System D	12	.1764	.0277
Within groups	45	.0697	.0015			System E	12	.1550	.0533
Total	47	.0733				System C	24	.1573	.0362
						Total	48	.1615	.0395

<u>Graphics tablet</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.1403	.0701	172.2518	.0000	System D*	12	.0297	.0211
Within groups	45	.0183	.0004			System E*+	12	.1344	.0343
Total	47	.1586				System C	24	.0032	.0047
						Total	48	.0426	.0581

[Appendix 26 continued]

Keyboard

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0009	.0004	1.5652	.2202	System D	12	.0156	.0168
Within groups	45	.0123	.0003			System E	12	.0132	.0176
Total	47	.0131				System C	24	.0227	.0159
						Total	48	.0185	.0167

Drawing Plan

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0004	.0002	.6177	.5437	System D	12	.0519	.0197
Within groups	45	.0158	.0004			System E	12	.0469	.0183
Total	47	.0162				System C	24	.0446	.0184
						Total	48	.0470	.0186

Speech list (hardcopy)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0000	.0000	.5721	.5684	System D	12	.0047	.0065
Within groups	45	.0017	.0000			System E	12	.0030	.0037
Total	47	.0018				System C	24	.0053	.0070
						Total	48	.0046	.0062

Effect of Integrated and Unitary Systems on Duration of Eye gaze to:

Graphics screen

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	147.2235	73.6118	.6715	.5160	System D	12	44.5892	8.7372
Within groups	45	4932.9294	109.6207			System E	12	41.8383	12.4089
Total	47	5080.1529				System C	24	40.3008	10.2138
						Total	48	41.7573	10.3966

[Appendix 26 continued]

Text screen

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	2148.4227	1074.2114	9.4604	.0004	System D	12	40.7017	7.4083
Within groups	45	5109.6768	113.5484			System E*	12	32.3225	14.7568
Total	47	7258.0995				System C	24	48.5092	9.5794
						Total	48	42.5106	12.4269

Graphics tablet

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	1873.8729	936.9365	147.4286	.0000	System D*	12	3.8800	2.6633
Within groups	45	285.9834	6.3552			System E*+	12	15.4425	4.3234
Total	47	2159.8563				System C	24	.2108	.3197
						Total	48	4.9360	6.7790

Keyboard

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	7.6137	3.8068	.8835	.4204	System D	12	1.8167	1.5183
Within groups	45	193.8884	4.3086			System E	12	1.7225	2.3433
Total	47	201.5021				System C	24	2.5633	2.1682
						Total	48	2.1665	2.0706

Drawing plan

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	14.5628	7.2814	1.1759	.3178	System D	12	7.2892	2.5229
Within groups	45	278.6465	6.1921			System E	12	6.5742	2.3855
Total	47	293.2092				System C	24	5.9529	2.5198
						Total	48	6.4423	2.4977

Speech list (hardcopy)

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.5775	.2888	.2264	.7983	System D	12	.7333	1.3732
Within groups	45	57.3857	1.2752			System E	12	.4575	.9560
Total	47	57.9632				System C	24	.6958	1.0752
						Total	48	.6456	1.1105

[Appendix 26 continued]

Effect of Integrated and Unitary Systems on Frequency of Hand:

<u>Idling</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.1076	.0538	100.0392	.0000	System D+	12	.0604	.0242
Within groups	45	.0242	.0005			System E*	12	.1609	.0334
Total	47	.1318				System C	24	.0481	.0155
						Total	48	.0794	.0529

<u>Drawing</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0003	.0002	1.2261	.3030	System D	12	.0302	.0107
Within groups	45	.0057	.0001			System E	12	.0352	.0108
Total	47	.0060				System C	24	.0291	.0117
						Total	48	.0309	.0113

<u>Searching menu item</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0031	.0016	51.3459	.0000	System D+	12	.0045	.0035
Within groups	45	.0014	.0000			System E*	12	.0196	.0106
Total	47	.0045				System C	24	.0000	.0000
						Total	48	.0060	.0098

<u>Selecting tablet command</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.1040	.0520	221.0346	.0000	System D+	12	.0115	.0128
Within groups	45	.0106	.0002			System E*	12	.1108	.0283
Total	47	.1146				System C	24	.0000	.0000
						Total	48	.0306	.0494

[Appendix 26 continued]

<u>Keying-in keyboard</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0003	.0002	1.5085	.2322	System D	12	.0061	.0113
Within groups	45	.0049	.0001			System E	12	.0040	.0053
Total	47	.0053				System C	24	.0101	.0118
						Total	48	.0076	.0106

Effect of Integrated and Unitary Systems on Duration of Hand:

<u>Idling</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	700.6073	350.3037	3.4670	.0398	System D	12	63.9817	5.8606
Within groups	45	4546.8068	101.0402			System E	12	62.2600	10.5590
Total	47	5247.4141				System C	24	70.6642	11.3110
						Total	48	66.8925	10.5663

<u>Drawing</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	490.5171	245.2586	3.4000	.0422	System D	12	30.3542	6.2555
Within groups	45	3246.0892	72.1353			System E	12	21.8792	8.7744
Total	47	3736.6063				System C	24	23.8887	9.2519
						Total	48	25.0027	8.9164

<u>Searching menu item</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	130.8065	65.4032	28.8444	.0000	System D+	12	1.0117	1.0214
Within groups	45	102.0353	2.2675			System E*	12	4.0283	2.8693
Total	47	232.8418				System C	24	.0000	.0000
						Total	48	1.2600	2.2258

[Appendix 26 continued]

Selecting tablet command

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	396.3419	198.1709	198.4327	.0000	System D+	12	.7717	.8244
Within groups	45	44.9406	.9987			System E*	12	6.8533	1.8455
Total	47	441.2825				System C	24	.0000	.0000
						Total	48	1.9063	3.0641

Keying-in keyboard

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	6.2077	3.1039	1.7905	.1786	System D	12	.8458	1.4479
Within groups	45	78.0077	1.7335			System E	12	.2317	.3210
Total	47	84.2154				System C	24	1.1121	1.5296
						Total	48	.8254	1.3386

Effect of Integrated and Unitary Systems on Frequency of word:

Recognition

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0343	.0172	21.7403	.0000	System D+	12	.0907	.0252
Within groups	45	.0355	.0008			System E*	12	.0387	.0264
Total	47	.0698				System C	24	.1036	.0301
						Total	48	.0841	.0385

Repetition

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0057	.0028	10.9489	.0001	System D+	12	.0333	.0151
Within groups	45	.0117	.0003			System E*	12	.0158	.0135
Total	47	.0174				System C	24	.0425	.0177
						Total	48	.0335	.0192

[Appendix 26 continued]

<u>Repeat due to Substitution</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0062	.0031	14.8873	.0000	System D*	12	.0196	.0118
Within groups	45	.0094	.0002			System E*	12	.0082	.0118
Total	47	.0156				System C	24	.0352	.0166
						Total	48	.0245	.0182

<u>Repeat due to Rejection/Spurious</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0004	.0002	2.3677	.1053	System D	12	.0163	.0114
Within groups	45	.0042	.0001			System E	12	.0089	.0064
Total	47	.0047				System C	24	.0156	.0102
						Total	48	.0141	.0100

<u>Repeat due to forgetting</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0003	.0002	6.0079	.0049	System D	12	.0047	.0078
Within groups	45	.0012	.0000			System E*	12	.0057	.0073
Total	47	.0016				System C	24	.0000	.0000
						Total	48	.0026	.0058

B. Performance Measures

Effect of Integrated and Unitary Systems on:

<u>Product quality</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	27.6875	13.8438	.6234	.5407	System D	12	7.0000	5.4439
Within groups	45	999.2917	22.2065			System E	12	8.6667	5.1932
Total	47	1026.9792				System C	24	6.8750	4.0466
						Total	48	7.3542	4.6745

[Appendix 26 continued]

<u>Production cost</u> (time)

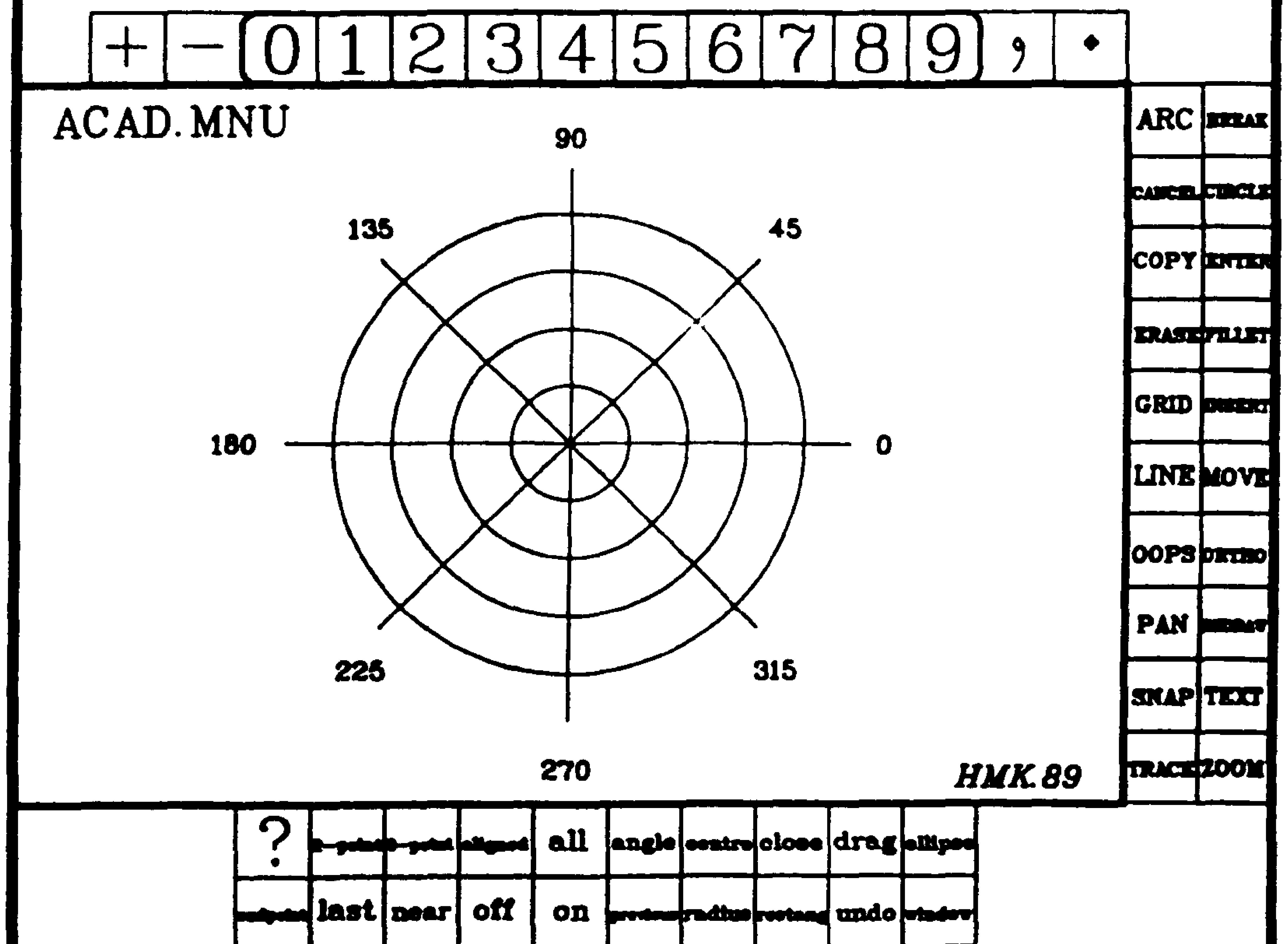
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	236.0975	118.0488	.6191	.5429	System D	12	44.0428	19.1452
Within groups	45	8579.9090	190.6646			System E	12	37.9937	10.8968
Total	47	8816.0065				System C	24	42.1926	11.8723
						Total	48	41.6054	13.6958

<u>Production cost</u> (efficiency)

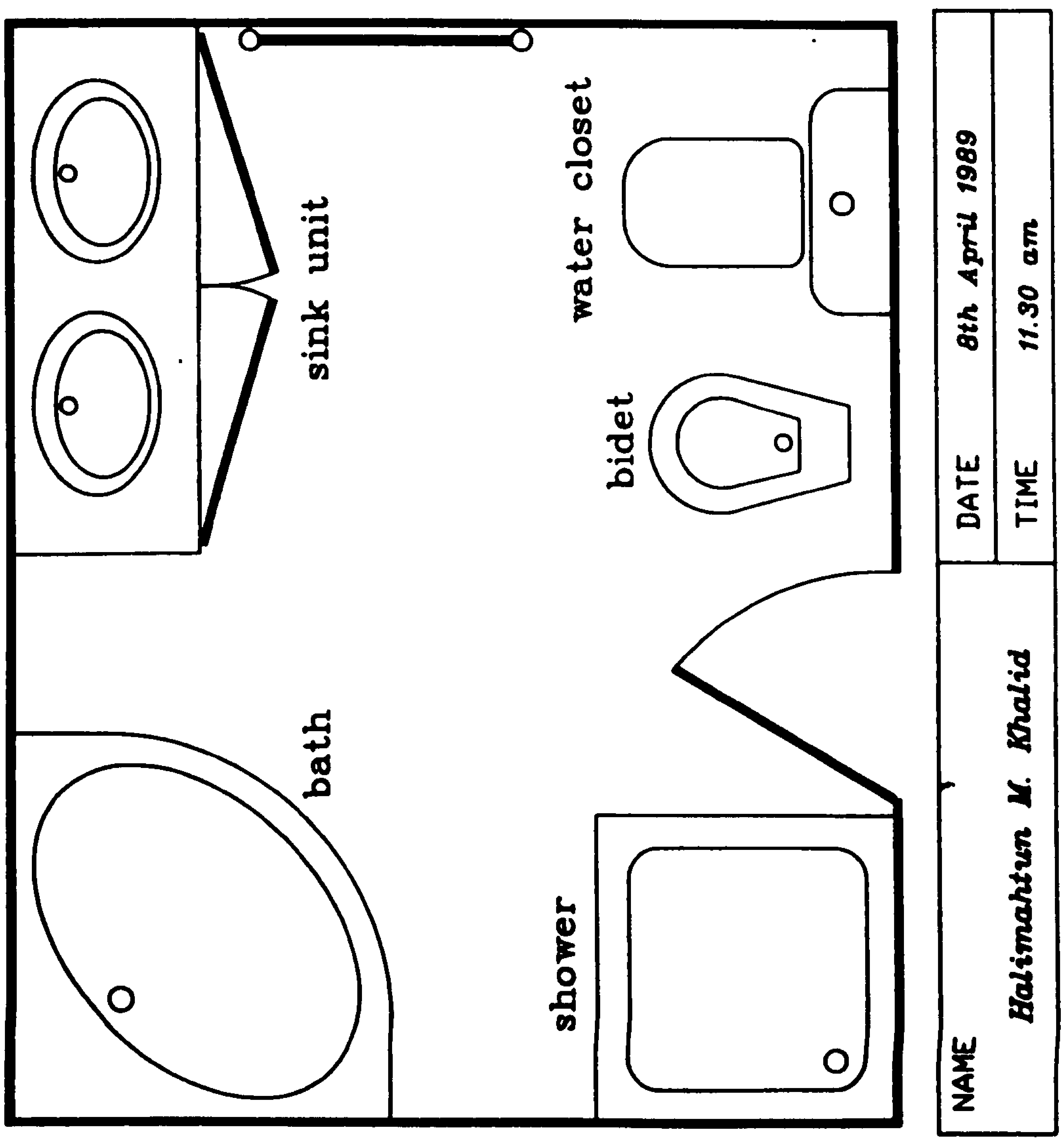
Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	1.1452	.5726	2.7309	.0760	System D	12	.9819	.3405
Within groups	45	9.4354	.2097			System E	12	1.3332	.7346
Total	47	10.5806				System C	24	.9739	.3110
						Total	48	1.0657	.4745

<u>Key:</u>
Results of Scheffe´ Test -
+ Integrated systems differed from each other
* Integrated systems differed from Unitary Speech

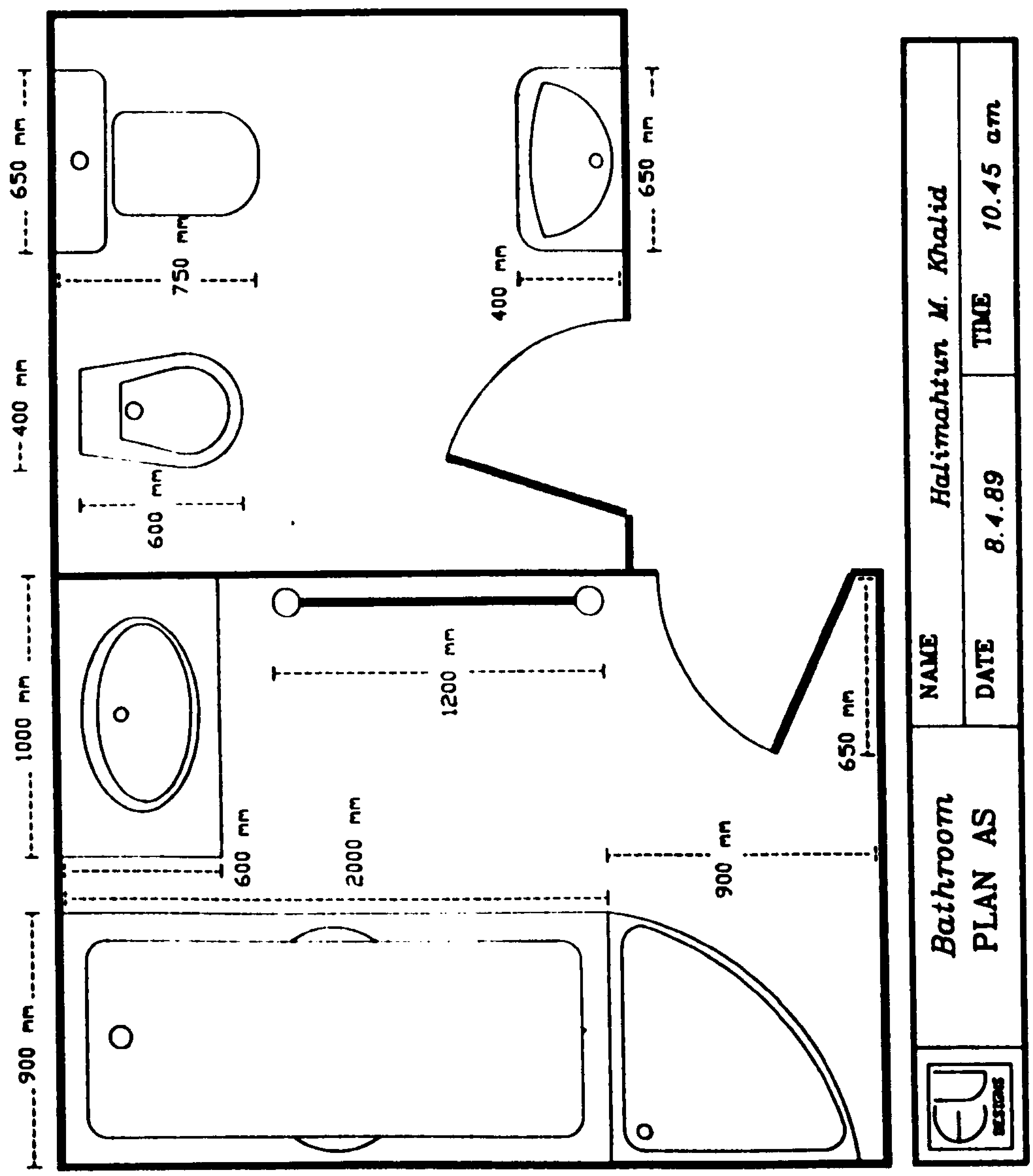
APPENDIX 27. Tablet Menu Overlay for Experiment 3



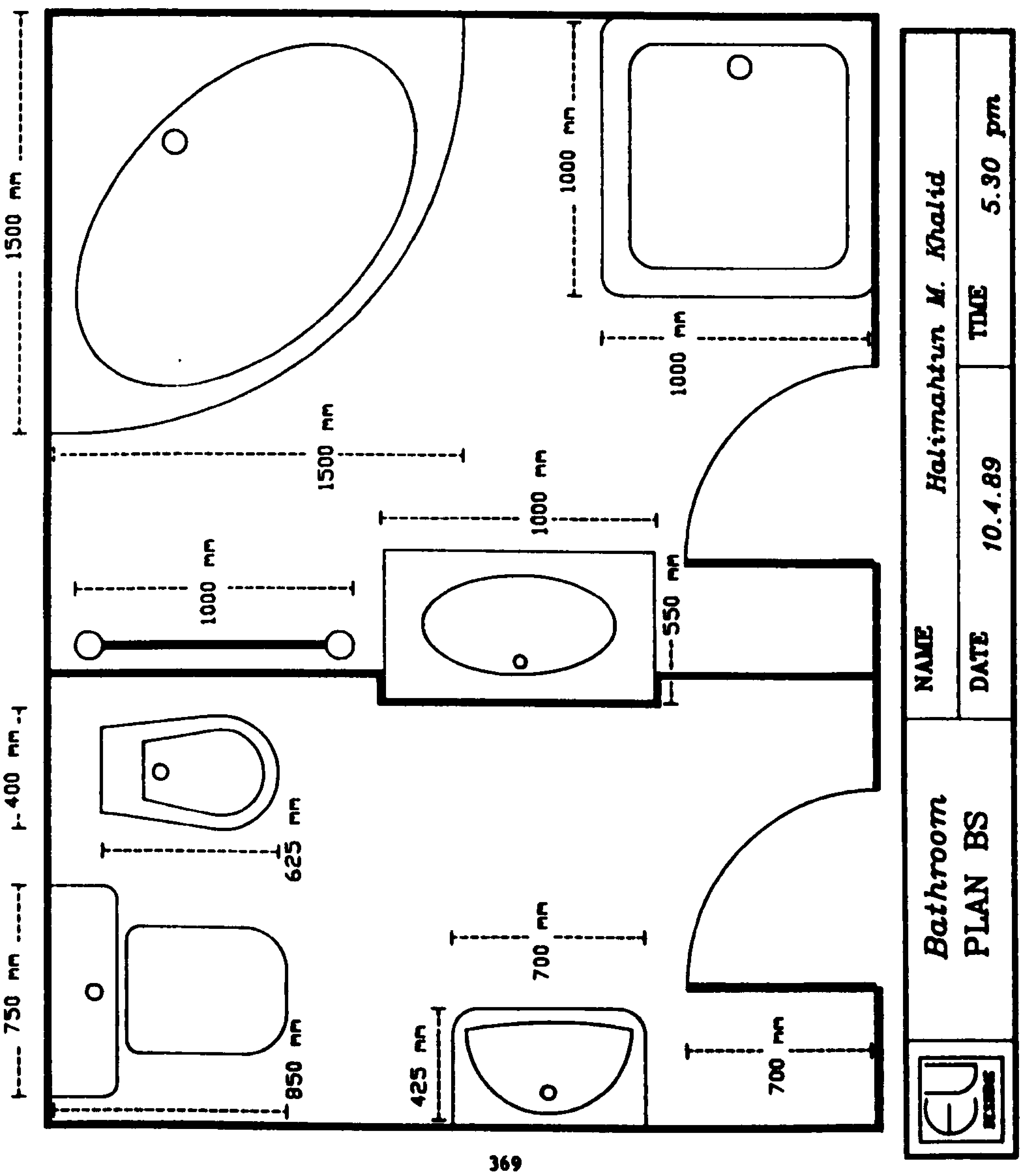
APPENDIX 28a. Training Task for Experiment 3



APPENDIX 28b. Experimental Task for Experiment 3 (First plan)



APPENDIX 28b. Experimental Task for Experiment 3 (Second plan)



APPENDIX 29. Speech List for Experiment 3

Check Speech List

Name.. Date..............................

	Word	*Confused with*	*Not recognised*
1	PAN	....	
2	REDRAW	....	
3	ZOOM	....	
4	all	....	
5	window	....	
6	ARC	....	
7	BREAK	....	
8	CIRCLE	....	
9	INSERT	....	
10	LINE	....	
11	angle	....	
12	2-point	....	
13	centre	....	
14	ellipse	....	
15	close	....	
16	undo	....	
17	3-point	....	
18	endpoint	....	
19	rectang	....	
20	CANCEL	....	
21	?	....	
22	LISTEN	....	
23	ENTER	....	
24	ERASE	....	
25	last	....	
26	COPY	....	
27	drag	....	
28	MOVE	....	
29	HELLO	....	
30	OOPS	....	
31	RETURN	....	
32	GOODBYE	....	
33	GRID	....	
34	on	....	
35	SNAP	....	
36	TRACE	....	
37	radius	....	
38	FILLET	....	
39	off	....	
40	SWITCHOFF	....	
41	TEXT	....	
42	aligned	....	
43	previous	....	
44	ORTHO	....	
45	near	....	
46	help	....	

APPENDIX 30a. Training Overview and Instructions for Experiment 3

1. INTRODUCTION

You are participating in an experiment which attempts to explore the use of speech and manual input devices for performing computer-aided design (CAD) tasks. This experiment is a part of my PhD research project. I am particularly interested in observing how you use these devices to carry out simple design tasks. Your actions will be video-filmed to enable analysis of device usage. So, this is NOT a test of your performance nor skills.

You will first be trained on the application software, that is, AutoCAD version 2.17. The aim of this brief training is to show you how to design and manipulate objects on a computer screen in order that you may achieve the basic skills required for performing **three** simple design tasks. Once trained you should be able to do the tasks on your own. Therefore, it is very important that you pay careful attention to the demonstration. Please ask questions whenever in doubt.

2. SESSIONS

There are **two** separate sessions. Today is your first session. You need to return for your second session within **1-2 days**.

Session 1 : *Training on Speech Recogniser and CAD*

<u>Phase 1</u>
* briefing on systems used
* complete subject profile form
* train voice and speech vocabulary
* demonstrate AutoCAD functions

<u>Phase 2</u>
* practise Task 1 for **45 minutes** using speech recogniser

Session 2 : *Perform two more tasks*

<u>Phase 3</u>
* perform Task 2
* complete short questionnaire

<u>Phase 4</u>
* perform Task 3
* complete short questionnaire

3. GENERAL INSTRUCTIONS

* Try to do as much as you possibly can and as **exactly** as shown in the plan.
* Try doing Tasks 2 and 3 without any help from me.
* Try to be **consistent** in the way you speak or verbalise a command word.

I may request you to stop at some point even if you have not finished drawing.
I hope you will enjoy learning AutoCAD using speech and tablet input.

Any questions?

APPENDIX 30b. Recording Form for Experiment 3

- -

Subject Name: Subject No:

- -

Training session : Date .... April/May 1989

		Start time	*End time*
1.	Briefing + Profile form	....	....
2.	Voice + Speech Vocab.	....	....
3.	CAD demo	....	....
4.	Practice	....	....

a. PLAN T

b. Time : **45 minutes**

- -

Comments:

==

Drawing Information [PLAN T]

1. Entities :

Final	....
Original	....
Difference	....

2. Errors : ....

Experimental Session: First task Date .... April/May 1989

- -

Condition	**First................Second...................**
System	**F [TM/TS]**
Plan	**A B**

- -

		Start time/no.	*End time/no.*
1.	Practice : YES / NO	....	....
2.	Perform : **35 minutes**	....	....
3.	Tape ID	....	....

Problems :

1. Speech
2. CAD
3. Others

[Appendix 30b continued]

Drawing Information

1. Entities

Final	Plan A [103]......................	Plan B [105].....................	
Original	Plan A : 34	Plan B : 36	
Difference	....	....	

2. Errors ....

Experimental session: Second task

- -

Condition	**First................**	**Second...................**
System	**G [SM/TS]**	**H [TM/GS+TS]**
Plan	**A**	**B**

- -

	Start time/no.	*End time/no.*
1. Practice : YES / NO	....	....
2. Perform : **35 minutes**	....	....
3. Tape ID	....	....

Problems :

1. Speech
2. CAD
3. Others

- -

Drawing Information

1. Entities

Final	Plan A [103]......................	Plan B [105].....................
Original	Plan A : 34	Plan B : 36
Difference	....	....

2. Errors..

APPENDIX 31. Specific Instructions for Experiment 3

Specific Instructions: System F and System G

In this session you will perform TWO more tasks

1. For BOTH tasks you will use the

    - speech recogniser to enter commands
    - graphics tablet to enter numerics and coordinates
    - keyboard to enter text

2. For the FIRST task, you will have

    - all systems information on the text screen
    - backup commands on the graphics screen

3. You must use the backup facility whenever your spoken command is not recognised. BUT you must not use it continuously. Its main function is to help you re-enter commands that are not recognised when spoken for the first time, so as to avoid repeating the command.

4. Be efficient and accurate in your drawing.

Specific Instructions: Second task

1. For this SECOND task, you will have

    - all systems information on the text screen
    - backup speech commands on the graphics tablet

2. Please use the backup facility whenever your spoken command is not recognised. You may check for prompts from the text screen.

3. Be efficient and accurate in your drawing.

[Appendix 31 continued]

Specific Instructions: System F and System H

In this session you will perform TWO more tasks

1. For BOTH tasks you will use the
   - speech recogniser to enter commands
   - graphics tablet to enter numerics and coordinates
   - keyboard to enter text
2. For the FIRST task, you will have
   - all systems information on the text screen
   - backup speech commands on the graphics tablet
3. You must use the backup facility whenever your spoken command is not recognised. But you must NOT use it continuously. Its main function is to help you re-enter commands that are not recognised when spoken for the first time, so as to avoid repeating the command.
4. Be efficient and accurate in your drawing.

Specific Instructions: Second task

1. For this SECOND task, you will have
   - all systems information on the text screen
   - systems prompts on BOTH graphics and text screens
   - backup commands on the graphics tablet
2. Please use the backup facility whenever your spoken command is not recognised. You may check for prompts from either screen.
3. Be efficient and accurate in your drawing.

APPENDIX 32a. Questionnaire for Experiment 3: System G

System G group [Questions for System F, see Appendix 32b]

Instructions:

1. Please answer all questions.
2. Place a slash mark on the line provided for Q1,Q2 and Q3.
3. Place a tick against the relevant response for the remaining questions, and where required, please explain. If in doubt, please ask.

A. Performance

Q1. How would you rate your performance for this task?

Q2. Are you satisfied with your performance?

Q3. How would you rate this system (in particular the input devices, and display screens) that you use to do the task?

poor I -- I excellent

Q4. What difficulties or problems did you experience in using this system (specifically the input devices and display screens) to carry out the task?

Q5. Given the problems above, how did you overcome in order to perform the task?

B. Preference

Q6. Did you find the backup commands for speech input useful?

Yes No Why? Please explain

Q7. Did you find having all systems information on one screen (ie. text screen) satisfactory.

Yes No Why? Please explain

Q8. Of the two systems you have used today, which do you prefer most. Please rank 1 for the most preferred system and rank 2 for least preferred.

System for Task 1 System for Task 2

Please explain why you prefer this system most?

Q9. Of the two modes for backup speech commands, which mode do you most prefer?

backup commands on screen
backup commands on tablet
neither mode (Go to Qn. 10)

Q10. Would you prefer a different backup mode altogether? Yes No

If YES, what sort of mode? Any idea? If NO, why?

Q11. Would you prefer to have systems information on both screens (ie. text and graphics screen)? Yes (go to Q12.) No (Go to Q13)

Q12. If YES, what kind of information and where? Please consider carefully.

prompts	TS......	GS.....
speech recognition feedback	TS......	GS.....
error feedback	TS	GS......
text input feedback	TS	GS......

Q13. If NO, why not? Please explain.

APPENDIX 32b. Questionnaire for Experiment 3 (System F/H)

Subject name:.. Date:.............. System: F / H

QUESTIONNAIRE

Instructions

1. Please answer ALL questions.
2. Place a slash mark (/) on the lines provided for Q1, Q2, Q3, and Q20.
3. Place a tick (/) against the relevant responses for the remaining questions, and where required, please explain.
4. If in doubt, please ask. Thank you for completing this questionnaire.

A. Performance

Q1. How would you rate your performance for this task?

|--|

excellent poor

Q2. Are you satisfied with your performance?

|--|

dissatisfied very satisfied

Q3. How would you rate this system (in particular the input devices and display screens) that you used to do the task?

|--|

poor excellent

Q4. What difficulties or problems did you experience in using this system to carry out the task?

Q5. Given the problems above, how did you overcome in order to perform the task?

Q6. In trying to complete this task, what was an important factor to you?

Be efficient (ie. accurate and quick) ....
Be accurate, not necessarily quick ....
Be quick, not necessarily accurate ....

B. System Preference

Q7. Of the TWO systems you have used today, which do you prefer most. Please give a rank 1 for the most preferred system and rank 2 for least preferred.

System 1 for Task 2 ....
System 2 for Task 3 ....

Please explain why you prefer the system you ranked 1 most.

Q8. Did you find the backup commands for speech input useful?

Yes .... No ....Why? Please explain.

[Appendix 32b continued]

Q9. Would you prefer a different backup mode to tablet for speech commands?

Yes ....[Go to Q10]   No .... [Go to Q11]

Q10. If YES to Q9., any of the following?

backup commands on screen ....
backup commands via keyboard ....
backup commands on puck ....
other [please suggest] ....

Q11. If NO to Q9., please explain why you prefer the tablet mode as a backup for speech input.

Q12. Did you find having all systems information on one screen (ie. text screen) satisfactory?

Yes ....   No ....

Why? Please explain.

Q13. Did you find having systems prompts on BOTH screens (ie. text and graphics screens) helpful?

Yes ....   No ....

Why? Please explain.

Q14. Would you prefer to have systems information on BOTH screens, besides the prompts?

Yes .... [Go to Q15]   No .... [Go to Q16]

Q15. If YES to Q14, what kind of information AND where? Please consider your suggestions carefully.

systems prompt	text screen	graphics screen ....
speech recognition feedback	text screen	graphics screen ....
error feedback	text screen	graphics screen ....
text input feedback	text screen	graphics screen ....

Q16. If NO to Q14., why not? Please explain.

C. Task and Training

Q17. Of the two tasks performed today, was one task more difficult than the other OR both were equally difficult?

one more difficult .... [Go to Q18]

both equally difficult .... [Go to Q19]

[Appendix 32b continued]

Q18. Which task was more difficult?

task 2 (ie. performed first) ....

task 3 (ie. performed second) ....

Q19. Was the training on AutoCAD sufficient to allow you perform the tasks?

Yes .... No ....

D. Speech recognition and Feedback

Q20. For this session, did you find using speech recogniser to perform the task....

a.

|--|

not at all tiring — very tiring

b.

|--|

quick — slow

c.

|--|

boring — enjoyable

d.

|--|

confusing — not confusing

e.

|--|

easy — difficult

21. In this session, you were provided with backup commands for speech input BUT you still seem to repeat some commands, why?

22. Do you need to check whether your spoken command is recognised?

Yes, all the time ....Yes, sometimes .... No ....

23. Do you often look at the systems prompts?

Yes, all the time ....Yes, sometimes .... No ....
[Go to Q24.]

If YES and SOMETIMES, which screen?

graphics screen more than text screen ....

text screen more than graphics screen ....

other [please clarify] ....

24. Any comment(s) to add on any aspects of the system (eg. input devices, display screen, task, etc.)? Please elaborate.

Thank you for participating in this experiment.

APPENDIX 33. ANOVA Summary - Experiment 3: New Hybrid Speech-Manual versus Old Hybrid Speech-Manual Systems (n=16)

A. BEHAVIOUR MEASURES:

Effect of New and Old Hybrid Systems on Frequency of Eye gaze to:

<u>All targets</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.2963	.1482	.7436	.4842	System G	8	2.5737	.3536
Within groups	29	5.7784	.1993			System H	8	2.3020	.3617
Total	31	6.0748				System F	16	2.4496	.5156
						Total	32	2.4437	.4427

<u>Graphics screen</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0134	.0067	4.6361	.0179	System G	8	.1847	.0203
Within groups	29	.0418	.0014			System H	8	.2254	.0538
Total	31	.0552				System F	16	.1760	.0353
						Total	32	.1905	.0422

<u>Text screen</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0068	.0034	2.2900	.1193	System G	8	.1466	.0246
Within groups	29	.0433	.0015			System H	8	.1191	.0488
Total	31	.0501				System F	16	.1546	.0386
						Total	32	.1437	.0402

<u>Graphics tablet</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0042	.0021	10.2925	.0004	System G*	8	.0088	.0052
Within groups	29	.0060	.0002			System H+	8	.0332	.0161
Total	31	.0102				System F	16	.0362	.0163
						Total	32	.0286	.0181

[Appendix 33 continued]

<u>Keyboard</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0000	.0000	.7680	.4731	System G	8	.0002	.0005
Within groups	29	.0001	.0000			System H	8	.0010	.0024
Total	31	.0001				System F	16	.0004	.0011
						Total	32	.0005	.0014

<u>Drawing plan</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0011	.0006	1.1029	.3454	System G	8	.0400	.0285
Within groups	29	.0145	.0005			System H	8	.0565	.0275
Total	31	.0156				System F	16	.0495	.0154
						Total	32	.0489	.0224

<u>Command list</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0000	.0000	.0228	.9775	System G	8	.0036	.0066
Within groups	29	.0009	.0000			System H	8	.0031	.0038
Total	31	.0009				System F	16	.0032	.0058
						Total	32	.0033	.0054

Effect of New and Old Hybrid Systems on Duration of Eye gaze to:

<u>Graphics screen</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	684.7013	342.3506	6.1558	.0059	System G	8	62.0263	6.7736
Within groups	29	1612.8095	55.6141			System H*	8	65.2038	9.2238
Total	31	2297.5107				System F	16	54.6406	6.8122
						Total	32	59.1278	8.6089

[Appendix 33 continued]

Text screen

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	465.7283	232.8642	4.1314	.0264	System G	8	27.0062	5.3304
Within groups	29	1634.5845	56.3650			System H*	8	19.6650	9.7334
Total	31	2100.3128				System F	16	28.9275	7.1764
						Total	32	26.1316	8.2312

Graphics tablet

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	131.3242	65.6621	9.3730	.0007	System G	8	1.4038	.7358
Within groups	29	203.1572	7.0054			System H	8	5.0150	2.8416
Total	31	334.4813				System F	16	6.3550	3.0859
						Total	32	4.7822	3.2848

Keyboard

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0127	.0063	.1715	.8432	System G*	8	.0425	.1202
Within groups	29	1.0711	.0369			System H+	8	.0988	.2559
Total	31	1.0838				System F	16	.0719	.1847
						Total	32	.0713	.1870

Drawing plan

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	4.1934	2.0967	.0776	.9255	System G	8	7.3425	5.4781
Within groups	29	783.6878	27.0237			System H	8	8.0963	4.6659
Total	31	787.8812				System F	16	8.2094	5.2992
						Total	32	7.9644	5.0414

Command list

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.7911	.3956	.0822	.9213	System G	8	.9688	1.9296
Within groups	29	139.5010	4.8104			System H	8	.7437	1.0502
Total	31	140.2921				System F	16	1.1275	2.6548
						Total	32	.9919	2.1273

[Appendix 33 continued]

Effect of New and Old Hybrid Systems on Frequency of Hand:

<u>Idling</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0001	.0000	.0476	.9536	System G	8	.0672	.0155
Within groups	29	.0200	.0007			System H	8	.0678	.0296
Total	31	.0200				System F	16	.0703	.0285
						Total	32	.0689	.0254

<u>Drawing</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0001	.0000	.2753	.7613	System G	8	.0372	.0070
Within groups	29	.0044	.0002			System H	8	.0410	.0151
Total	31	.0045				System F	16	.0373	.0129
						Total	32	.0382	.0121

<u>Locating menu item</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0005	.0003	1.8551	.1745	System G	8	.0258	.0143
Within groups	29	.0040	.0001			System H	8	.0145	.0095
Total	31	.0045				System F	16	.0204	.0112
						Total	32	.0203	.0120

<u>Entering command</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0002	.0001	.2587	.7738	System G	8	.0229	.0145
Within groups	29	.0084	.0003			System H	8	.0206	.0183
Total	31	.0086				System F	16	.0257	.0175
						Total	32	.0237	.0166

[Appendix 33 continued]

<u>Entering numeric</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	.0000	.0000	.5566	.5792	System G	8	.0043	.0036
Within groups	29	.0004	.0000			System H	8	.0031	.0039
Total	31	.0004				System F	16	.0049	.0039
						Total	32	.0043	.0038

Effect of New and Old Hybrid Systems on Duration of Hand:

<u>Idling</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	3.5079	1.7540	.0624	.9397	System G	8	58.9000	6.0656
Within groups	29	815.7832	28.1305			System H	8	57.9838	6.3471
Total	31	819.2911				System F	16	58.5788	4.2914
						Total	32	58.5103	5.1409

<u>Drawing</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	97.2452	48.6226	1.4539	.2502	System G	8	32.4913	6.2073
Within groups	29	969.8532	33.4432			System H	8	37.4013	7.6326
Total	31	1067.0984				System F	16	35.2650	4.4147
						Total	32	35.1056	5.8671

<u>Locating menu item</u>

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	51.6588	25.8294	5.2472	.0114	System G*+	8	5.5787	3.4662
Within groups	29	142.7524	4.9225			System H	8	2.3525	2.0142
Total	31	194.4112				System F	16	2.8462	1.4202
						Total	32	3.4059	2.5043

[Appendix 33 continued]

Entering command

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	1.3524	.6762	.6497	.5297	System G	8	1.6250	1.0006
Within groups	29	30.1850	1.0409			System H	8	1.1225	.9995
Total	31	31.5374				System F	16	1.5806	1.0387
						Total	32	1.4772	1.0086

Entering numeric

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.4739	.2370	.7038	.5029	System G	8	.7787	.4871
Within groups	29	9.7639	.3367			System H	8	.4600	.5569
Total	31	10.2378				System F	16	.7112	.6289
						Total	32	.6653	.5747

Effect of New and Old Hybrid Systems on Frequency of Word:

Recognition

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.0004	.0002	.1874	.8301	System G	8	.0867	.0292
Within groups	29	.0330	.0011			System H	8	.0961	.0333
Total	31	.0334				System F	16	.0884	.0359
						Total	32	.0899	.0328

Repeat due to recogniser errors

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.0001	.0000	.2362	.7912	System G	8	.0308	.0130
Within groups	29	.0050	.0002			System H	8	.0277	.0102
Total	31	.0051				System F	16	.0315	.0143
						Total	32	.0304	.0128

[Appendix 33 continued]

Repeat due to user errors

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.0000	.0000	.0871	.9168	System G	8	.0111	.0074
Within groups	29	.0020	.0001			System H	8	.0121	.0081
Total	31	.0020				System F	16	.0126	.0087
						Total	32	.0121	.0080

B. PERFORMANCE MEASURES:

Effect of New and Old Hybrid Systems on:

Product quality

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	4.6875	2.3438	.0616	.9404	System G	8	9.1250	4.5178
Within groups	29	1102.8125	38.0280			System H	8	9.7500	9.7358
Total	31	1107.5000				System F	16	8.8125	4.4455
						Total	32	9.1250	5.9771

Production cost (time)

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	21.3213	10.6606	.0417	.9592	System G	8	44.5429	14.8981
Within groups	29	7419.5357	255.8461			System H	8	46.8423	19.0426
Total	31	7440.8570				System F	16	45.8388	14.8941
						Total	32	45.7657	15.4928

Production cost (efficiency)

Source	D.F.	Sum of Squares	Mean Squares	F ratio	F prob.	System	No.	Mean	Std. dev.
Between groups	2	.0319	.0159	.0772	.9259	System G	8	1.7021	.4808
Within groups	29	5.9837	.2063			System H	8	1.7904	.4259
Total	31	6.0156				System F	16	1.7553	.4543
						Total	32	1.7508	.4405

[Appendix 38 continued]

<u>User acceptability</u> (performance rating)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	101.5655	50.7827	.1652	.8485	System G	8	59.1848	13.3290
Within groups	29	8914.8748	307.4095			System H	8	58.9674	23.4540
Total	31	9016.4402				System F	16	62.6359	15.9595
						Total	32	60.8560	17.0544

<u>User acceptability</u> (satisfaction rating)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	948.1451	474.0725	1.1843	.3203	System G	8	50.2174	16.0003
Within groups	29	11608.7547	400.3019			System H	8	65.5978	22.3775
Total	31	12556.8998				System F	16	58.3967	20.5125
						Total	32	58.1522	20.1261

<u>User acceptability</u> (recogniser rating)

Source	*D.F.*	*Sum of Squares*	*Mean Squares*	*F ratio*	*F prob.*	*System*	*No.*	*Mean*	*Std. dev.*
Between groups	2	4.8263	2.4132	.0055	.9945	System G	8	59.6196	19.4710
Within groups	29	12764.1186	440.1420			System H	8	60.2717	25.1573
Total	31	12768.9449				System F	16	60.5707	19.4594
						Total	32	60.2582	20.2954

<u>Key:</u>

Results of Scheffe´ Test :
+ new hybrids differed from each other
* new hybrids differed from old hybrid.

APPENDIX 34. Proposed Initial Set of Guidelines

GUIDELINES FOR INTEGRATING SPEECH AND MANUAL INPUT IN CAD SYSTEMS

Introduction

This document describes the development and application of guidelines for the combined use of speech and manual input in CAD systems. The successful application of speech and manual input in human-computer interfaces depends not only upon technological advances in speech and/or manual input hardware and software, but also on the development of empirically based human factors guidelines. The term *guidelines*, as used here, refers to a set of recommendations or suggestions. Hence, they are intended to be prescriptive but without being able to offer guarantees.

Computers today are used for a broad range of applications. User interface design guidelines cannot be applied usefully in every case (Smith & Mosier, 1986). Some computer systems are designed to help particular users perform specific tasks. For example, CAD systems are designed to support designers in the design process. To the extent that CAD systems support users performing CAD tasks, careful design of the user-computer interface is crucial to ensure effective interaction as well as to optimise design behaviour. The guidelines proposed here are intended to improve user interface design of CAD systems that integrate speech and manual input in a single use.

Integrated Speech-Manual Systems

While great advances in speech recognition technology may be envisaged, there still remains the need to understand and maximise the effectiveness of speech interfaces using current technology (Starr, 1987). The thesis investigations have demonstrated that the unitary use of speech or manual input devices to perform CAD tasks is non-optimal. In particular, their use incurred costs to the user. Integrating speech and manual input within a single system has proven to result in more optimal design behaviour and task performance. Therefore, the guidelines proposed here are intended to support the configuration and design of speech plus manual input devices in CAD applications.

The need for design guidelines in the development and refinement of input devices is much emphasised in the human factors literature (eg. Nickerson, 1986; Buxton, 1986; Davis & Swezey, 1983). There is general agreement that most of the design guidelines that have been suggested lack empirical support (Meister, 1988; Nickerson, 1986). The guidelines proposed here are not based on judgement, rather they use the experimental data (ie. the experimental findings and users' comments) to generate a partial set of recommendations as a guide for users of interactive systems. As such, they are <u>not</u> complete and potential users would still need to refer to other handbooks on guidelines for advice on aspects not covered here.

Importance of User Requirements

Users of CAD systems interact with a computer in order to accomplish design tasks. The users tend to differ in ability, training, experience and attitudes. Design of speech-manual interface must take into account these human factors. Failure to incorporate user requirements in the design of improved systems will result in low user acceptability. The thesis has identified pertinent problems experienced by users in using the demonstrator CAD system. These problems emerged as user complaints, indicating considerable effort on the users' part to adapt to the configured system and to develop short-term coping strategies so as to overcome the problems. The guidelines proposed here take account of user requirements expressed by users of the demonstrator system.

In short, these guidelines are potential solutions to the problems documented in the thesis investigations concerning the use of unitary speech or unitary manual input. The solutions are needed in order to optimise design behaviour and task performance.

[Appendix 34 continued]

Guidelines Organisation

The guidelines are formal to the extent that they are explicitly expressed. To use the guidelines effectively, they must be translated into specific design rules. This is because a guideline can have different possible translations to users. It is also possible that a particular guideline (eg. concerning design of speech vocabulary) could conflict with another guideline (eg. concerning allocation of device functions to data type). Failure to translate the guidelines into highly prescriptive design rules can result in inconsistent design. "It is only specifically worded design rules that can be enforced, not guidelines" (Smith & Mosier, 1986, p. 9). In other words, these guidelines are offered to users as a potential resource, rather than as a contractual design standard.

The format for expressing the guidelines follows the standard format by Smith and Mosier (1986). The guidelines are organised in a single section; within the section, the guidelines are grouped by specific functions. Under any function, there will be guidelines pertaining to related subordinate topics. The section begins with an introductory discussion of design issues relating to data entry and information display. The discussion provides some perspective for the guidelines that follow. The guidelines themselves are numbered sequentially in order to permit convenient referencing. Each guideline has been given a short title to indicate its particular subject matter. Following its number and title, each guideline is stated as a single sentence. Guidelines are worded as simply as possible.

A stated guideline will be illustrated by one or more examples. The examples are derived from the investigations. It is important to emphasise that examples are presented here only to illustrate and are not intended to limit the interpretation of guidelines. Examples are followed by comments. The comments are clarifications of a guideline to provide the reasoning behind a guideline using material from the thesis investigations. Where a comment is related in some way to other published reports, references will be made citing author(s) and date. The references are listed in the bibliography section of the thesis.

Guidelines Application

The proposed guidelines may be used in two ways: (1) to configure a speech-plus-manual interface for CAD; and (2) to improve the design of a speech and/or manual interface for CAD systems. The guidelines are intended for use by three classes of users: (1) end users; (2) implementors of CAD systems; and (3) system designers. However, it should be pointed out that the guidelines are not developed with the third group in mind. Therefore, it is not clear whether they can actually apply them in designing novel integrated systems. With regard to the first and second groups of users, the guidelines may be tailored to meet their requirements in configuring speech-manual CAD systems, particularly using existing systems. The aim is to optimise the use of available systems, adding speech to the interface, in order to develop a speech-plus-manual input system. Thus, providing a solution to CAD users and implementors in making good use of their available systems rather than to invest in new CAD and/or speech systems.

It should also be emphasised that the proposed guidelines are intended for CAD applications. In other words, it is not clear whether the guidelines can be applied to other applications such as word processing, database management, etc. However, with appropriate modifications, and providing there is sufficient documented evidence that the rules have been consistently applied within an application, the design rules derived from these guidelines might later be used for other applications.

As stated above, before a guideline can be applied it must be translated into specific design rules. For example, a guideline which states that system information should be distributed in dual-screen displays might be translated into design rules that specify where various information types should appear, such as prompts on graphics screen while command entries on text screen, etc.

[Appendix 34 continued]

DATA ENTRY AND INFORMATION DISPLAY

Data entry refers to user actions involving input of data to a computer, and computer responses to such inputs (Smith & Mosier, 1986). *Information display* refers to computer output of information to a user. Here, displayed information is concerned with providing guidance to a user in performing a CAD task.

Data can be entered into a computer in a variety of ways. In CAD, users might designate position or direction by pointing at a display using input devices such as the graphics tablet, mouse or keyboard. Users might enter numeric and/or textual data by keyed or spoken inputs; and users might draw pictures or manipulate displayed graphic elements with the tablet. Rationalising the various functions to be supported by the input devices should help to make the system more simple and easy to use. The thesis has clearly shown the importance of identifying data configuration and device functionality through analyses of system behaviour and task performance, and applying this as a basis for integrating speech and manual input in a single system.

The main objective of integrating speech and manual input is to optimise behaviour and performance. Additionally, to improve data entry functions through minimal input actions, minimal memory load on the user, and flexibility of user control of data entry.

The computer also plays a role in the data entry process, guiding users who need help, checking data entries to detect errors, and providing other kinds of data processing aids. Such user guidance includes prompts, error messages, advisory messages, status information to help guide a user's interaction with a computer. The fundamental objectives of information display are to promote efficient system use with minimal memory load on the user and with flexibility for supporting users of different skill levels.

Thus, flexibility is an important concept in system design. The specific means of achieving such flexibility must be spelled out in design guidelines (Smith & Mosier, 1986). The thesis investigation experimentally compared different configurations of input modes to specific aspects of task (ie. data type). By providing alternatives to the default speech data entry mode, users had the flexible choice of deciding the input mode for command entry, which in turn influenced their acceptance of the system. Flexibility is also needed so that users can tailor information displays online to meet their requirements.

Objectives:
Optimal behaviour and performance
Minimal entry actions by user
Minimal memory load on user
Flexibility for user control of data entry and display

1.0 *Device functions and data allocation*

1 Clearly defined functions for input devices
The functions of each input device should be well rationalised and clearly distinct from each other.
Example: Assign speech input to command entry while keyboard to text entry.
Comment: Clearly defining the functions that each device supports helps to optimise the utility of each input device and to simplify its use. This view is supported by Monk (1986) and Whitefield (1986a). Thus, devices should be assigned function(s) to which it is best suited. There is evidence to suggest that the keyboard is suited to text entry, the tablet for graphical entry, the recogniser for command entry.

[Appendix 34 continued]

2 Flexible assignment of input devices to data type

Allocation of input mode to data should be flexible so that it does not load user memory.

Example: Keep commands together and allocate commands to one input mode.

Comment: Inflexible allocation of different input devices to data type can increase memory load on the users' part and tends to be potentially difficult for user learning. Dividing commands and assigning them to separate input modes caused users to verbalise some commands that were not available within the particular input mode. Thus, commands should be combined as a whole and not divided between input modes.

3 Commands distinctly separate from numerics

Ensure that commands are kept separate from numeric data by assigning each data set a different input mode.

Example: Assign commands to speech input but numerics to tablet input.

Comment: Current speech recognition devices are not well developed and tend to be error prone. Because of speech confusability, user inconsistency and time per utterance length, commands were generally found to be confused with sub-commands and numbers, and vice versa. Since numerics are fewer in number and tended to contain fewer characters (<4) than most commands, their separation from commands might reduce confusability.

1.1 *Speech input*

4 Single function for speech input

Assign speech input to perform a single function to which it is best suited.

Example: Use speech input to enter commands only or coordinates only but not both.

Comment: Due to poor performance of current speech recognisers, speech input should be used to perform a single function. Thus, there should be some good reason for choosing the device function to which speech input will be assigned.

5 Backup facility for speech input

When speech input is used as a default data entry mode, provide facilities to which a user can fall back when a spoken entry is not recognised.

Example: A backup facility can be in the form of a retraining facility or a menu.

Comment: Speech recognition requires some form of backup facilities to enable data reentry. The facility should be simple to use so that the costs of use would be much less than the cost of re-verbalising the entry. The need for a backup facility is also emphasised by Waterworth and Talbot (1987), Hapeshi and Jones (1988).

6 Performance aids for speech input

Provide performance aids to ease memory load when using speech input.

Example: A pop-up screen menu or an online speech list can be used to aid user memory.

Comment: Speech input involves recall from memory and in the event of high speech confusability, the occurrence of memory failure is thereby increased. Using a hardcopy speech list as a memory aid for speech input was detrimental to performance. Thus, performance aids should be in the form of an online facility that would increase on-screen gazing. The need for performance aids to support task is much discussed by Bailey (1982).

7 Limited vocabulary for speech input

Structure the vocabulary used for speech entry such that the vocabulary size is manageable by the user.

Example: Design the speech vocabulary according to task requirements, consisting mainly of frequently-used or important words.

Comment: In view of speech confusability and user memory, it is crucial that the vocabulary for spoken data entry is limited to some options. To increase the likelihood that a user's valid entries are correctly identified by the system, the user's vocabulary should be predictable. A vocabulary is predictable when a user's choice of inputs at any given time is small, so that the system will be more likely to make a correct match in interpreting an entry. Criteria such as frequency of use and importance of use for the task could be applied in deciding which entry is to be spoken. This then will reduce the size of the vocabulary. It was established that naive users and novices were able to remember between 50-75 per cent of the vocabulary (ie. about 50-70 words). This familiarity criterion could also be used as a basis for structuring the vocabulary.

8 Templates for vocabulary

Provide more than one template for each item in the speech vocabulary.

Example: Use two or more templates per word.

Comment: In view of the high number of word repeats due to speech confusability, enrolment templates should be increased irrespective of the recogniser type. Connected speech recognisers that employ a single pass training procedure are not robust enough to cope with current speech problems. Thus, increasing the number of templates per word, as often suggested in the speech literature (eg. Hapeshi & Jones, 1988; McCauley, 1984; Talbot, 1987), might resolve this.

9 Alternative entries for speech input

When speech input is the primary mode for data entry, provide alternatives for critical entries so that if the system cannot recognise an entry another entry can be substituted.

Example: "Return" might be defined as an acceptable substitute for "Enter".

Comment: Because speech recognition is affected by normal variations in a user's "task" voice, by changes in the acoustic environment and by displacement of the microphone, etc., a spoken entry that was accepted previously might not be accepted at a later time. Thus, for important entries a user should be able to use an alternative word that is acoustically-dissimilar and preferably not monosyllabic.

10 Phonetically distinct vocabulary for speech input

Ensure that the spoken entries for any transaction are phonetically distinct from one another.

Example: Avoid using words that are similar-sounding such as "Enter" and "centre", particularly if the entry has critical consequences for the task.

Comment: Words which are easily confused by the speech recogniser should be replaced with other dissimilar-sounding words. Thus, words should be tested and checked in order to determine which the device tends to confuse, and which words it can distinguish reliably. This view is also expressed by Smith and Mosier (1986), Waterworth and Talbot (1987).

11 Alternative method for speech input activation

Design alternative methods to spoken entries for switching on and off the speech recogniser.

Example: Designate a key on the keyboard for this purpose or a control button near the screen display.

Comment: Because of user inconsistency, using only speech commands to activate and deactivate the speech device has been shown to induce frustration and anxiety on the part of the user. Switching off the device during critical aspects of task performance is important for some users, but failing to do so because of poor recognition increases confusability. The significance of having this alternative facility was also made by Martin (1989).

12 PAUSE and CONTINUE options for speech input

Provide PAUSE and CONTINUE options for speech input, so that a user can stop speaking without having to log off the recogniser.

Example: A user may wish to stop speaking data for a time in order to speak to a colleague or to answer a telephone.

Comment: Because the users were inexperienced in CAD, the need to request help verbally was inevitable. To do so users were forced to log off the device each time they wish to say something that is not intended as an entry. Thus, speech recognition devices should be provided with facilities (eg. a manual switch) that would enable them to be kept on-hold while the user switches to another conversational task. This suggestion is similar to that

[Appendix 34 continued]

made by Smith and Mosier (1986).

13 Easy error correction for speech input
Provide simple error correction procedures for speech input, so that when a spoken entry has not been correctly recognised, the user can cancel that entry and speak again.
Example: The use of explicit CANCEL action that is not tied to other task functions might be one way of interrupting the execution of the verbalised input.
Comment: The need for some form of error correction procedures is widely discussed in the speech literature (eg. Williges, Schurick, Spine & Hakkinen, 1986; McCauley, 1984; Hapeshi & Jones, 1988) Error correction procedures independent of CAD functions are needed to support speech input. It has been demonstrated that users employed the CANCEL command much more than any other command in order to terminate the execution of a confused command. Because CANCEL sometimes is confused with another word, the tendency to revert to keyboard use becomes inevitable. This in turn resulted in increased eye and hand transitions, thereby incurring costs.

14 Improvement of speech input reception
Design the recogniser such that the input reception is not constrained by the hardware.
Example: Provide a control mechanism on the headset microphone that can improve speech input reception.
Comment: Current hardware of speech recognisers presents different problems to different users. A headset microphone tended to constrain head movements because of the trailing wire which connects it to the computer. Because the positioning of the microphone is crucial to good input reception, and the fact that users are not able to maintain the same microphone position throughout its use, speech input reception has been affected. Also, the pressure from the use of a headset microphone caused discomfort to users. Thus, system designers should consider some alternatives to current speech hardware components.

15 Improvement of speech input training
Provide alternative ways to the normal list-reading technique of training the speech recogniser.
Example: Randomising the order in which the word list is read during training or using phrases, etc. are some ways of improving speech template training.
Comment: Reading aloud a list of single words several times, as is usually the case, can be monotonous and unnatural. This is reflected in the users' intonation: they will speak in a dull, flat voice which is not reproduced during the execution of the actual task. Varying the order in which the word is read during training should help to remove the predictability of the words such that the rhythmic patterning inherent in reading lists of words would be reduced. Additionally, with connected speech recognisers, the use of phrases rather than single words might help to remove the 'unnaturalness' and monotony related to speech input training.

1.2 *Manual input*

16 Minimal use of manual data entry
Data entry via manual input mode should be kept to a minimum so that a user can stay with one manual method of entry, and not have to shift to another.
Example: Minimise the use of two input devices that require the same output resources so as to reduce hand transitions from tablet to keyboard and vice versa.
Comment: Shifting of one hand between two input devices sharing the same modality (such as the keyboard and tablet) has been shown to incur behavioural costs. For users who are not able to touch type or key-in tablet items without visual aid, the need to gaze away from the primary display to manipulate the input devices occurs frequently, besides incurring time. This problem was also documented by Van der Heiden and Grandjean (1984), as well as Monk (1986).

[Appendix 34 continued]

17 Limited design vocabulary for tablet input

Structure the design vocabulary for tablet input according to task requirements so that it speeds up visual search and is manageable to the user.

Example: Limit the number and type of items in the tablet overlay or screen menu to reduce visual load.

Comment: The use of tablet and/or screen menus requires visual search. Having too much information in the menu will slow down search and incurs more time in menu selection. For naive users, this was shown to be detrimental to task performance. Thus, design items per menu should be of a reasonable size, and selective in accordance with task needs.

18 Organisation of tablet and screen menus

Design tablet and screen menus such that the items are not cluttered nor overloading on user memory.

Example: Arrange items in the menu according to some task-relevant criteria such as command set, frequency of use, etc.

Comment: Organisation of menu items is crucial to menu selection process. Items should be arranged following some logical grouping principles which do not conflict with each other nor with task. Examples of such principles are given in the literature on guidelines (eg. Cole, Lansdale & Christie, 1987; Bailey, 1982; McKenzie, 1988; Davis & Swezey, 1983). Arrangement of tablet items in alphabetical order was detrimental to visual search but this did not affect screen menu items. Hierarchical arrangement may be suitable for screen menu.

19 Responsiveness to transducer pen-down

The tablet should remain unresponsive to slight pressure from the transducer as would occur if the user swept the stylus or puck slightly over the tablet surface.

Example: Dragging the stylus over the tablet during drawing should not be registered as an input.

Comment: Sensitivity of the tablet to slight pressure has been a problem to some users. Coupled with stylus sensitivity, this has led to numerous data reentries. Because of varying skill level and work style, users are not able to use the transducer in the optimal position suggested. Thus, design of the tablet and the probes are crucial to avoid unnecessary inputs. This problem was also raised by Davis and Swezey (1983).

20 Improvement of tablet-transducer

Design the transducer for the tablet such that it does not interfere with menu selection and pointing processes nor constrain hand movements.

Example: The wire connecting the stylus to the computer can be problematic as it tends to trail hand movement.

Comment: Current transducers such as the puck and stylus all have wires that tend to obstruct user's pointing actions. There is a tendency to hold the wire with one hand while the other hand makes the selection, thus keeping both hands frequently busy. There is also a tendency for users to hold the transducer in the hand while the hand shifts to the keyboard. Thus, the trailing wire tends to obstruct key-pressing activity. When its use is combined with speech input, the wire tends to become entangled with the wire from the headset microphone.

1.3 *Information display*

21 Flexible allocation of system information to displays

Allocate system information flexibly between displays in dual-screen configuration so that users have the choice of which display to view for the information.

Example: Assign prompts and command feedback on the graphics and text screens, while other forms of feedback are assigned to the text screen to enable users process information selectively from the displays.

Comment: System information (such as prompts) that is needed by one group of users (eg. novices) should be displayed on both screens. But other users (eg. experienced) should be provided with alternatives to by-pass the standard user guidance facilities. Having prompts on both screens has helped to increase graphics screen gazing in CAD, besides reducing between-screen eye transitions.

22 Indicating recognition status of speech input

Provide some indication of speech recognition status to users as a feedback mechanism.

Example: Computer-generated speech in the form of prompts might be used to provide status messages.

Comment: Status information on speech input is particularly needed because of unreliable speech recognition. There should be a good reason to choose suitable feedback mechanisms that do not interfere with the task. Thus, the use of visual and/or auditory feedback must be considered carefully. If status information is not provided by design, users will be forced to gaze at the display that provides information on command feedback for acknowledgement. This behaviour incurs time and is thus non-optimal.

23 Feedback for control entries

Provide some indication of information processing status whenever the complete response to a user entry will be delayed, particularly for speech input.

Example: Display time-to-completion or some other indication of progress.

Comment: Indicating the progress of computer processing is particularly important with speech input. This is to avoid users from making another entry, thinking that the previous entry was not recognised. Also, users may be able to perform other tasks while waiting.

24 Error messages for verbal repeats

Display a non-disruptive error message if a user repeats an entry that is already recognised.

Example: Display the same entry but with changing annotation, perhaps marked with an asterisk to indicate a repeat.

Comment: If a spoken entry is repeated because the user may be uncertain whether it was recognised the first time round, displaying a brief error message should help to alert the user. The display should be timed so as to minimise disruption of the user's thought process and task performance.

APPENDIX 35a. Guidelines development process

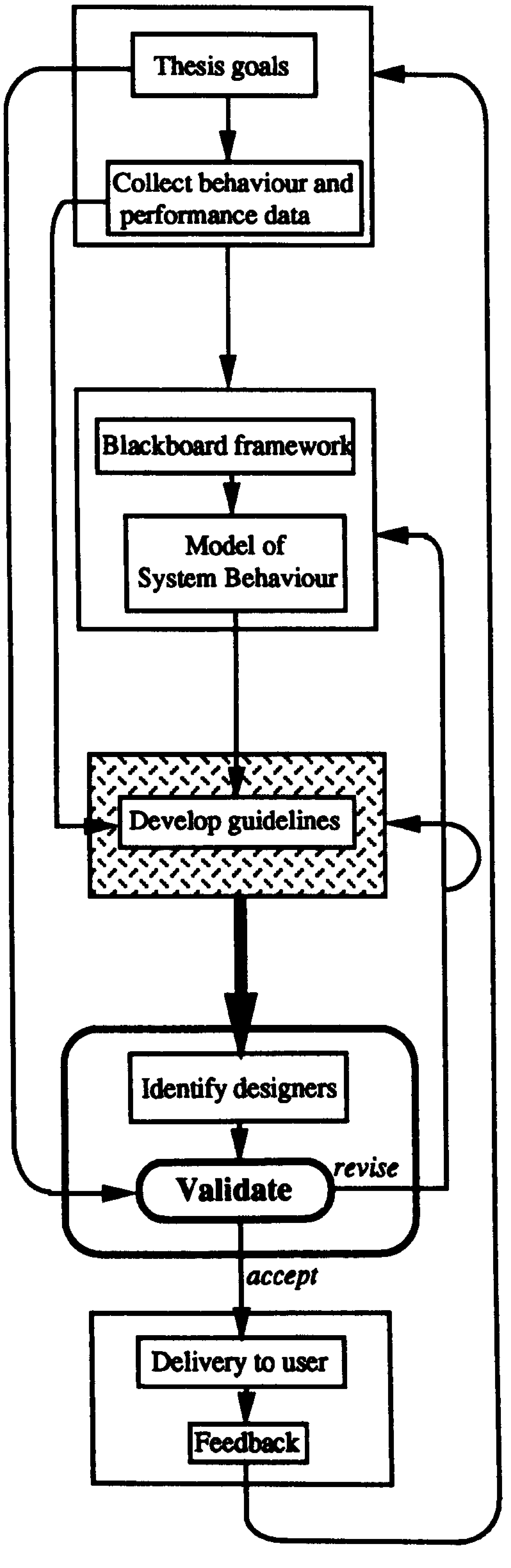


Figure showing the stages of developing, validating and refining the guidelines to achieve a usable design tool

APPENDIX 35b. Interview Schedule for Validation of Guidelines

Approach to Validation of Guidelines

1. **Background information**

    - Age
    - Job title
    - Time in present job
    - Previous job, time
    - Education/Design experience
    - Application for which you design systems
    - Type of system (hardware/software/etc.)

2. **Current design practice**

    - How would you set about designing a system (eg. a speech input system/CAD software/etc.) ?
    - What difficulties do you encounter?
    - Do you use Human Factors guidelines in designing systems?
        - If YES, which/what guidelines? How do you use them?
        - If NO, why not?
    - Do you use <u>other</u> guidelines?
        - If YES, which/what guidelines?

3. **Proposed guidelines**

    - Do you find each guideline
        - clear?
        - useful?

4. **Usability of proposed guidelines in system design/configuration**

    - Would you be able to use the guidelines?
    - Do you think other intended users could use the guidelines?

APPENDIX 36. Suggestions by System Designers for Revising the Guidelines

1 Clearly defined functions for input devices

The functions of each input device should be well rationalised and clearly distinct from each other.

- Qualify the example, that is, speech recognisers have a potential to be best suited (...) [D3]
- Modify the statement: "the type of input should be well rationalised", not the functions of each input device so as to achieve the degree of flexibility that is needed. [D4]
- Indicate in the comment that there are conflicting guidelines. Give explicit examples of that conflict and how to resolve them. [D1, D6]
- Distinguish this guideline clearly from guidelines 2 and 3. [D7]
- Explain further why certain input devices are suited to the functions mentioned. [D2]

2 Flexible assignment of input devices to data type

Allocation of input mode to data should be flexible so that it does not load user memory.

- Define the terms *user, input mode* and *flexibility.* [D1]
- Qualify the example, that is, "do not split commands between input modes". [D1, D2, D3]
- Explain in the comment that a flexible interface is one whereby all the input devices can do everything. [D2, D4]
- State the criteria for allocating input modes to data type. [D5]
- Specify what goes where so that designers do not have the options to make a choice. [D4]

3 Commands distinctly separate from numerics

Ensure that commands are kept separate from numeric data by assigning each data set a different input mode.

- Keep together the reasoning about poor speech performance within guideline 1 to avoid contradictory comments. [D3]
- Drop the word "might" if it has been proven that commands should be kept separate from numerics. [D4]
- Clarify the terms *command* and *subcommand,* that is, are they part of the same subset or to be treated separately. [D5]
- Suggest that numeric entry is quicker using the keypad on the keyboard. [D2]

4 Single function for speech input

Assign speech input to perform a single function to which it is best suited.

- Explain what constitutes a good reason? [D3]
- Explain why speech input should have a single function. [D2]
- Define the term *coordinates;* are they tablet coordinates or numerics? [D6]
- Combine this guideline with guideline 3 since the difference is not clear. [D4]

5 Backup facility for speech input

When speech input is used as a default entry mode, provide facilities to which a user can fall back when a spoken entry is not recognised.

- Define the term *retraining* facility. [D4]
- Include other backup facilities such as repair and recovery loops. [D5]
- Clarify that too much retraining can result in a set of quite diverse and unrepresentative templates. [D7]

6 Performance aids for speech input

Provide performance aids to ease memory load when using speech input.

- Define the terms *performance aids* and *online speech list.* [D4]
- Explain what detrimental effect(s) a hardcopy speech list has on performance. [D3]
- Give examples of aids that could trigger a pop-up menu, etc. such as snapping the finger, simple utterances like "Err..". [D1]
- Suggest some graphical aids (icons, etc.) that could help memory. [D2]

[Appendix 36 continued]

7 Limited vocabulary size for speech input
Structure the vocabulary used for speech entry such that the vocabulary size is manageable by the user.
• Replace the word *structure* with *design*. [D1, D2]
• Indicate an optimum vocabulary size. [D3]
• If known, suggest words that are best suited to commands and how to associate spoken commands with their functions. [D3]
• Clarify in the comment that if the words used are different from the basic CAD commands, it may confuse the user who is trained on the basics. Unless CAD system is designed with speech interface in mind, this problem holds. [D2]
• Identify the criteria on which to base decisions when designing the vocabulary. [D3]
• User memory may be one criterion, naturalness of words may be another. [D6]

8 Templates for vocabulary
Provide more than one template for each item in the speech vocabulary.
• Clarify that this depends on the recogniser, but generally it is better to use more than one template per word. [D7]
• Elaborate the example by suggesting: to store multiple templates as separate items or to average them together to make a composite template for each word; and to automatically adapting the templates to changes in the user's speech. [D7]
• Explain the benefits of multiple templates per word, that is, it saves retraining time and pays off in the long run. [D2]

9 Alternative entries for speech input
When speech input is the primary mode for data entry, provide alternatives for critical entries so that if the system cannot recognise an entry another entry can be substituted.
• Suggest alternative entries that do not vary in terms of their intended meaning, such as "polyline" and "pline" to aid remembering. [D2]
• Suggest the use of combined commands rather than single command, such as "zoom window", etc. to discriminate between alternatives. [D2]
• Specify the criteria for selecting alternative entries. [D3]
• If "monosyllabic" is a valid issue, state it as an individual guideline. [D3]
• Make this guideline consistent with previous related guidelines. [D4]

11 Alternative method for speech input activation
Design alternative methods to spoken entries for switching on and off the speech recogniser.
• Provide another example for activating or deactivating the speech input, that is, by turning on/off a button on the microphone. [D4]
• Suggest in the example a footswitch or something that is in any case mediated through a different channel to both the speech and manual entries. [D7]
• Combine this guideline with guideline 12 as they are both closely related. [D4]
• Emphasise the need for alternative methods to increase flexibility of use. [D2]

13 Easy error correction for speech input
Provide simple error correction procedures for speech input, so that when a spoken entry has not been correctly recognised, the user can cancel that entry and speak again.
• Indicate the input mode for this procedure, whether via the keyboard, tablet or speech. Avoid using speech input for implementing this procedure as it may generate more problems. [D3]
• Suggest other methods such as error repair and recovery loops. [D5]
• Suggest a CANCEL command outside of the speech input mode such as a two-position footswitch, one position for a toggled on/off and the other to cancel an entry. [D7]
• Suggest a button on the transducer itself to serve as a cancellation function to speech input. [D2]

[Appendix 36 continued]

14 Improvement of speech input reception

Design the recogniser such that the input reception is not constrained by the hardware.

• Suggest the use of a clip microphone because the wire trails from the body rather than from the head. [D2]

• Suggest a wireless headset microphone. [D1, D4]

• Explain that users should be provided with choices of speech facilities. [D6]

15 Improvement of speech input training

Provide alternative ways to the normal list-reading technique of training the speech recogniser.

• Suggest a task-like training. That is, collect the speech samples while the user is talking through a dummy run of the task. [D7]

• Suggest a simulated task or a practice task from which speech samples are collected rather than the normal recogniser training. [D2]

• Explain that the example needs further research since it is not clear that randomising the word order is suitable. [D5, D6]

16 Minimal use of manual data entry

Data entry via manual input mode should be kept to a minimum so that a user can stay with one manual method of entry, and not have to shift to another.

• Define the term *costs* and how they are measured. [D3]

• Clarify what is meant by *minimal entry* and *modality*. [D1]

• Provide a higher level, overriding design principle which would resolve the conflict between minimal entry and shifting of hand between input devices, given that both are in the same statement. [D6]

17 Limited design vocabulary for tablet input

Structure the design vocabulary for tablet input according to task requirements so that it speeds up visual search and is manageable to the user.

• Change the title to reflect the recommendation. [D4]

• Explain in the comment that with practice and experience, visual search may not necessarily be a problem. [D1]

18 Organisation of tablet and screen menu

Design tablet and screen menus such that the items are not cluttered nor overloading on user memory.

• Indicate the optimum number of items in the menu. [D4]

• Explain what is meant by *cluttered* and*hierarchical arrangement*. Emphasise that hierarchical menus may not be easy to use. [D1,D2]

• Emphasise on the navigational aspects of menu selection rather than the structure of the menu itself. [D6]

19 Responsiveness to transducer pen-down

The tablet should remain unresponsive to slight pressure from the transducer as would occur if the user swept the stylus or puck lightly over the tablet surface.

• Define what constitutes "slight pressure"? [D3,D5]

• Qualify the comment by suggesting to designers to try out with users and if they failed to register a response *n* times, use this as a criterion. [D6]

• Change the title as it is not clear. [D4]

• Explain that this problem is peculiar to a stylus rather than a puck. Also, a puck is more stable than a mouse and stylus in menu/entity selection. [D2]

20 Improvement of tablet transducer

Design the transducer for the tablet such that it does not interfere with menu selection and pointing processes nor constrain hand movements.

• Change the title to "Minimise the number of wires". [D4]
• Suggest in the example to use wireless transducers. [D1, D2, D4]
• Suggest a non-contact infra-red link between the stylus/puck and computer. [D7]
• Suggest a puck that requires one to squeeze it in order to manipulate it rather than dragging it across the tablet. [D2]

21 Flexible allocation of system information to displays
Allocate system information flexibly between displays in dual-screen configuration so that users have the choice of which display to view for the information.
• Clarify the users in this guideline as they could be end users, system designers, etc. [D1]
• Make this guideline generic to single screen systems as well. For example, "dual-screen" may be seen as equivalent to "windows" on a single screen. [D4, D1]
• Emphasise the importance of providing alternatives for different user levels. [D4, D7]
• Qualify the comment of the guideline as it is not clear. [D1]
• Suggest the use of screen menus that can be moved about in different parts of the display(s) as desired to increase flexibility. [D2]
• Emphasise that the command line on the primary screen should be more than one line than that currently available in order to provide adequate feedback . [D2]

22 Indicating recognition status of speech input
Provide some indication of speech recognition status to users as a feedback mechanism.
• Qualify the example concerning computer-generated speech since there is no evidence yet that it could be used. [D3]
• Make a separate guideline on the use of speech synthesis for feedback using evidence in the speech literature. [D3]
• Explain the term *feedback mechanism* and the criteria for choosing the mechanisms. [D4, D1]
• Suggest other modes in the example, such as a symbol that changes colour when a speech input is recognised. [D1]
• Suggest a facility that draws user's attention to it, such as a flashing light, a distinct beep, etc. or a combination of visual and auditory. [D2]

23 Feedback for control entries
Provide some indication of information processing status whenever the complete response to a user entry will be delayed, particularly for speech input.
• Make explicit the specific nature of the facility for indicating progress, that is, to count down always. [D6]
• Provide examples like several dots on screen which get eliminated as the task nears completion. [D2]

24 Error messages for verbal repeats
Display a non-disruptive error message if a user repeats an entry that is already recognised.
• Avoid using an asterisk to indicate an error as it may be interpreted as something important. Use other forms of symbol (eg. !) instead. [D3]
• Provide an error message in the same channel as the error occurs. [D7]
• Modify the comment: the non-disruptive message is not an error message since it is not a user error. [D1]
• Indicate that there should be feedback on what the system actually recognises, whether the first entry or the repeat entry. [D2, D3]
• Clarify that the system should accept the repeat but the user must be warned of it and the system must be able to cancel the repeat if it is the same as the first entry. [D2]
• Clarify that this example has not been tested, thus it is a suggestion for further research. [D5]